

**Original citation:**

Denrell, Jerker, Liu, Chengwei and Le Mens, Gaël. (2017) When more selection is worse. *Strategy Science*, 2 (1). pp. 39-63.

Permanent WRAP URL:

<http://wrap.warwick.ac.uk/85716>

**Copyright and reuse:**

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions. Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

**Publisher's statement:**

<https://doi.org/10.1287/stsc.2017.0025>

**A note on versions:**

The version presented here may differ from the published version or, version of record, if you wish to cite this item you are advised to consult the publisher's version. Please see the 'permanent WRAP URL' above for details on accessing the published version and note that access may require a subscription.

For more information, please contact the WRAP Team at: [wrap@warwick.ac.uk](mailto:wrap@warwick.ac.uk)

# **When More Selection is Worse**

**Jerker Denrell**

Warwick Business School, Scarman Road, University of Warwick, Coventry, CV4  
7AL UK, denrell@wbs.ac.uk

**Chengwei Liu**

Warwick Business School, Scarman Road, University of Warwick, Coventry, CV4  
7AL UK, Chengwei.Liu@wbs.ac.uk

**Gaël Le Mens**

Department of Economics and Business, Universitat Pompeu Fabra, C/Ramn Trias  
Fargas, 25-27, 08005 Barcelona, SPAIN, gael.le-mens@upf.edu

## **Abstract**

We demonstrate a paradox of selection: the average level of skill among the survivors of selection may initially increase but eventually decrease. This result occurs in a simple model in which performance is not frequency dependent, there are no delayed effects, and skill is unrelated to risk-taking. The performance of an agent in any given period equals a skill component plus a noise term. We show that the average skill of survivors eventually decreases when the noise terms in consecutive periods are dependent and drawn from a distribution with a ‘long’ tail - a sub-class of heavy-tailed distribution. This result occurs because only agents with extremely high level of performance survive many periods and extreme performance is not diagnostic of high skill when the noise term is drawn from a long-tailed distribution.

**Keywords:** Organizational Evolution, Evolutionary Economics, Organizational Ecology

**Acknowledgments:** We are grateful for comments from Bill Barnett, Mike Hannan, Thorbjorn Knudsen, Dan Levinthal, Jim March, and Sidney Winter.

## 1. INTRODUCTION

Suppose you observe an industry with the purpose of identifying and imitating best practice. You know that firms do not change their capabilities much over time due to learning or forgetting. The population of firms changes over time as a result of entry and exit. Low performing firms tend to exit the industry whereas highly performing firms tend to stay in. When should you observe the industry? In the early stages of the industry? Later on, when many firms have exited? Or in the middle-stages? A standard evolutionary argument ('survival of the fittest') suggests that you should observe the firms remaining in the later stages of the industry as these firms have survived selection for a longer time and are likely to be the most capable. In this paper, we demonstrate that there are conditions under which learning from those who have survived for the longest possible time is suboptimal and that one would be better off learning from those who have survived for a shorter time.

Of course, it is well known that selection does not necessarily increase the proportion of the most capable firms – those that would have the highest performance in the long run, had they survived (Wright, 1931; Levins, 1968; Holland, 1975; Nelson & Winter, 1982, 2002). When the 'fitness' of a practice depends on how many others have adopted this practice or on the presence of specific organizational characteristics, an inferior practice could become dominant (Wright, 1931; Maynard Smith, 1982; Arthur, 1989; Carroll & Harrison, 1994; Levinthal, 1997). Similarly, when selection operates on short-term performance, it might eliminate practices with positive long-term effects, especially in changing environments or when firms can adapt (Levins, 1968; Elster, 1979; Nelson & Winter, 1982; Levinthal & March, 1981; Levinthal & Posen, 2007; Levinthal & Marino, 2015). Finally, selection can be biased against practices that lead to highly variable performance even if mean performance is high (Cohen, 1966; Denrell & March, 2001; Levinthal & Posen, 2007).

Here we identify another, in some sense more basic, reason for why firms that have survived for a longer time might not have capabilities superior to firms that have survived for a shorter time. We analyze a model in which we assume away all of the above complications. The key to our argument is that the relation between the organizational characteristics and selection is probabilistic. Selection operates on performance in the sense that the lowest performing firms exit the population while the others stay in. But performance is determined not only by organizational characteristics but also factors unrelated to organizational characteristics, i.e., luck. We show that when the association between organizational characteristics and actual performance is highly variable (i.e., luck plays a large role), it can happen that selection initially leads to an increase in the prevalence of superior characteristics and later to a *decrease* in the prevalence of superior characteristics! Such inefficient

selection happens when luck has persistent effects. In contrast, when luck has non-lasting effects (e.g., the contribution of luck to performance changes from period to period), selection is efficient and leads to an increased prevalence of superior characteristics.

In this paper we focus on just one aspect of evolutionary explanations – selection – and leave out many other important aspects. In particular, the dynamics of skills and routines, which is central to the evolutionary theory of the firm (Nelson & Winter, 1982), is not explored here. Nelson & Winter (1982) assume that profitable firms expand while unprofitable firms contract. Moreover, they assume that unprofitable firms are more likely than profitable firms to search for new routines. Their focus is on how routines evolve as a result of such market driven search processes (Winter, 1971). Our focus is instead on the process of elimination of relatively poorly performing units. Using the terminology of Hodgson & Knudsen (2010), we focus on ‘subset’ selection (selecting a subset of units for survival, without changing their properties) while the evolutionary theory of the firm focuses on ‘successor’ selection (in which the units being selected change through imperfect replication involving mutation but also, in social science applications, processes of learning and search).

Evolutionary theories and selection explanations are sometimes viewed as alternative forms of explanations distinct from theories that emphasize goal-directed and adaptive behavior by managers. We do not pursue this agenda here. Our focus is on the population level consequences of selection regardless of the level of intentionality and rationality of the agents involved. Most selection models in economics and management assume that agents are intentional and goal-directed (Nelson and Winter, 1982, p. 10-11). And while many evolutionary accounts assume that agents are boundedly rational instead of profit maximizing (Alchian, 1950; Winter, 1964) there are several well-known selection models in economics that assume that agents are rational. For example, selection in the model of Jovanovic (1982) (i.e., exit by a firm) occurs as the result of optimal stopping based on Bayesian updating.

In what follows, we first provide an illustration of our argument using simulations of a simple model. Then we clarify the positioning of our work within the literature on selection in strategy and organization theory. We describe our model and our formal analysis. We establish formally the conditions under which selection is efficient (i.e., selection leads to an increasing prevalence of superior characteristics over time) or inefficient (i.e., selection leads to a decreasing prevalence of superior characteristics over time). Then we show, using computer simulations, that our basic result holds when we relax some of the simplifying assumptions of our formal model. Finally, we discuss

when our technical results are relevant and discuss the implications of our findings for empirical phenomena such as organizational obsolescence.

## 2. ILLUSTRATION

To illustrate the basic idea, consider the following simple model. There are  $n = 50$  firms. The performance of firm  $i$  in period  $t$ ,  $p_{i,t}$ , is equal to the sum of its skill,  $u_i$ , and a noise term,  $\varepsilon_i$ :  $p_{i,t} = u_i + \varepsilon_i$ . There are two levels of skill: high ( $u_i = 1$ ), and low ( $u_i = 0$ ), each equally likely. The level of skill and the value of the noise term remain constant during the lifetime of a firm. It follows that the performance of firm  $i$  is the same in every period. This setup represents a situation when initial ‘luck’ has strong long-term effects.

Selection works as follows. In each period  $t$  the 10 percent lowest performing firms (based on  $p_{i,t}$ ) are removed from the population. Each eliminated firm is replaced by a new firm. Consider a new firm  $j$ . The probability that it has high skill ( $u_j = 1$ ) is 50% and its performance is  $p_{j,t} = u_j + \varepsilon_j$  where  $\varepsilon_j$  is drawn from the same density  $f_\varepsilon$ . Note that both  $u_j$  and  $\varepsilon_j$  remain constant during the lifetime of a firm. We are interested in how the proportion of high skill firms in the population evolves over time. Does it increase monotonically with time? Are older firms, that survived more selections, more likely to have high skill?

Figure 1A shows that when the noise terms are drawn from a normal distribution, the proportion of high skill firms increases over time. The pattern differs, however, when the noise terms are drawn from a Student’s t-distribution. As shown in Figure 1B, the proportion of high skill firms initially increases but eventually decreases and ultimately becomes close to 50%, the proportion of high skill firms at the beginning. What is noteworthy is that firms that have survived for a longer time (e.g., 100 periods) are *less* likely to have high skill ( $u_i = 1$ ) than firms that survived for a shorter time (e.g., 20 periods). The mechanism underlying the dynamics depicted on Figure 1B thus leads to ‘inefficient’ selection. Although firms with the highest level of performance are more likely to survive in every period, and firms with high skill have higher expected performance, selection based on performance does not increase the proportion of high skill firms. In contrast, after some time, it leads to a *decrease* in the proportion of high skill firms.

Why does the proportion of high skill firms decline over time when the noise term is drawn from a t-distribution? The precise mechanism is discussed in the formal analysis in Section 4 and the intuition is explained in more detail in Section 5. At this stage, note that (1) we know that survivors must have had a very high level of performance (because only firms with a very high level of performance survive many periods); and (2) it is known from prior research that when

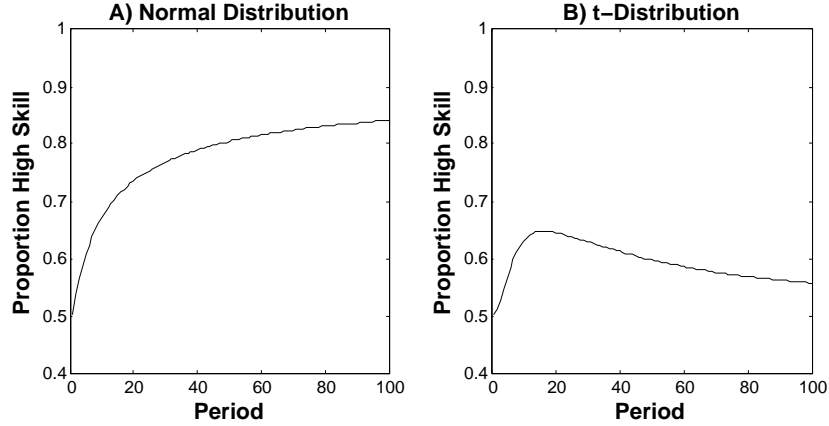


FIGURE 1. Proportion of high skill firms as a function of time. In each period the firms with the 10 percent lowest performances are removed and replaced with new firms. The noise terms remain the same across periods and are drawn from A) a normal distribution with mean 0 and variance 1 and B) a t-distribution with mean 0 and 1 degree of freedom. Each graph is based on 10,000 simulations, each with  $n = 50$  firms.

performance partly depends on luck and luck is drawn from a ‘long-tailed’ distribution, such as the t-distribution, an extremely high level performance is not diagnostic of high skill (Weibull et al., 2007; Denrell & Liu, 2012). Hence, having survived many periods can in fact be less diagnostic of high skill than surviving fewer periods. This is because surviving many periods requires an extreme level of performance, whereas surviving fewer periods requires a high but not extreme level of performance.

Importantly, this mechanism only operates when the performance in period  $t$  is strongly dependent on the performance in prior periods (conditional on skill level). In particular, it does not apply when the noise terms are redrawn in every period. Consider the following variation of the model. As before, the performance of firm  $i$  in period  $t$ ,  $p_{i,t}$ , equals its skill,  $u_i$  plus a noise term. But here, we assume that the noise terms are not drawn just once at the time of entry but are redrawn in every period. Thus, we have a similar specification, but with a time index on the noise term  $\varepsilon_{i,t}$ :  $p_{i,t} = u_i + \varepsilon_{i,t}$ . We assume that the noise terms are *iid* for all  $t$  and  $i$ . Figure 2 displays the evolution of the proportion of high skill firms for the two noise distributions used before (Normal with mean 0 and variance 1 and t-distribution with mean 0 and 1 degree of freedom). Here, the

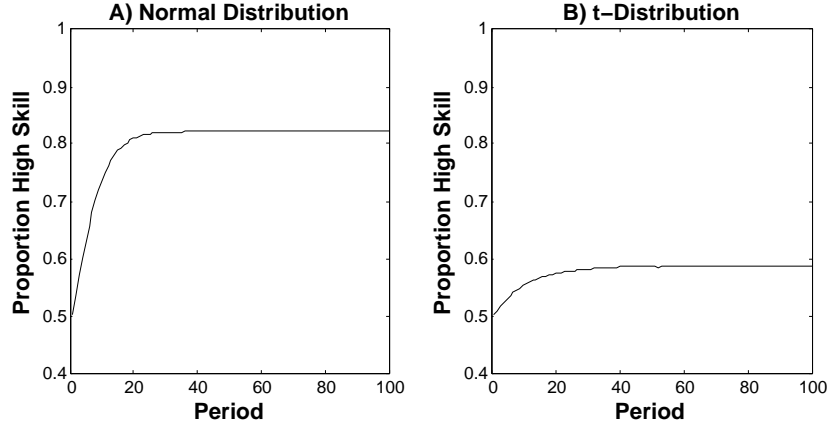


FIGURE 2. Proportion of high skill firms as a function of time. In each period the firms with the 10 percent lowest performances are removed and replaced with new firms. The noise terms are redrawn in every period from A) a normal distribution and B) a t-distribution. Each graph is based on 10,000 simulations, each with  $n = 50$  firms.

proportion of high skill firms monotonically increases over time in both cases. In contrast to what happened in the previous setting (Figure 1B), selection is efficient even when the noise term follows a t-distribution (Figure 2B).

Why is the proportion of high skill firms monotonically increasing when noise terms are independently drawn in every period but not when they are drawn just once at the time of entry (in the case of a t-distribution)? When the noise terms are independently drawn in every period survival until the end of period  $t$  requires that a firm passes  $t$  distinct ‘tests’. All these tests are independent *conditional on skill*. In this case, survival during early periods (the early ‘tests’) is informative about skill since survival during these early periods does not require extreme levels of performances. By contrast, when the noise term is drawn just at the time of entry, all  $t$  tests are dependent.

It is worth noting that inefficient selection can occur even when firm performance does not remain constant during the lifetime of a firm. For example, performance may follow a random walk (Levinthal, 1991; Denrell, 2004; Le Mens et al., 2011). Specifically, consider a firm that enters in period  $t$ . Its performance in period  $t$  follows the same specification as before:  $\varepsilon_{i,t}$ :  $p_{i,t} = u_i + \varepsilon_{i,t}$ . Its performance in period  $t + 1$  is  $p_{i,t}$  plus an independent draw from the same noise term distribution:  $p_{i,t+1} = u_i + \varepsilon_{i,t} + \varepsilon_{i,t+1}$ . Similarly,  $p_{i,t+2} = u_i + \varepsilon_{i,t} + \varepsilon_{i,t+1} + \varepsilon_{i,t+2}$ , etc. In this setup, performance

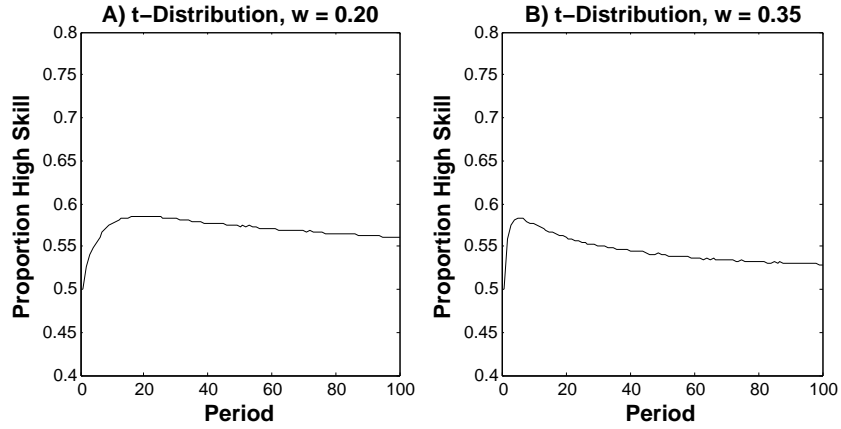


FIGURE 3. How the proportion of high skill firms changes over time when the noise term follows a random walk and A)  $\varepsilon_{i,t}$  are drawn from a t-distribution with 1 degree of freedom and  $w = 0.2$  and B)  $\varepsilon_{i,t}$  are drawn from a t-distribution with 1 degree of freedom and  $w = 0.35$ . Each graph is based on 10,000 simulations, each with 50 firms, where the  $w$  percent firms with the lowest performances are replaced in every period.

is not constant but changes in every period. Simulations show that if selection is strong enough ( $w$ , the proportion of firms that exit in each period, is high enough), the proportion of high skill firms will decline with time after some time (see Figure 3 for a depiction of the dynamics of this proportion for two levels of  $w$ ). More generally, in Sections 6 and 7 we show that our basic result holds under a number of alternative assumptions about selection, replacement, and performance dynamics.

### 3. PRIOR LITERATURE ON THE EFFICIENCY OF SELECTION

How do our results compare to prior work on selection in strategy and organization theory? Prior work has demonstrated that selection processes in markets and organization are limited in how effectively they remove ‘inferior’ organizations from a population. Some researchers have argued that selection processes may be only loosely coupled with performance (Meyer & Zucker, 1989). The threshold for exit can depend on alternative employment opportunities of an entrepreneur in addition to firm performance (Gimeno et al., 1997). Well-connected and powerful firms can sometimes survive for a long time even if they are inefficient (Perrow, 2002). Large firms can also



become buffered from selection forces (Levinthal, 1991; Barnett, 1997). Other researchers have demonstrated that even if selection processes reliably select on the basis of economic performance, the most efficient organizations may fail to survive. As Levinthal and Posen write: ‘Even if selection is effective in removing inferior organizations at one point in time, it may be ineffective over time in that it may remove organizations that, had they survived, would have gone on to do well.’ (Levinthal and Posen, 2007, p. 587). Two well-known mechanisms leading to such inefficient selection are time-dependent fitness and frequency-dependent fitness.

Consider time-dependent fitness. Fitness may change over time because firms or their environments change, but selection is often myopic and responds only to current levels of fitness (Elster, 1979; Levinthal & March, 1981; Levinthal & Posen, 2007). Such myopic selection may eliminate units with high future potential but low current performance. As Nelson and Winter note: ‘[i]f firms are small in the early stages of industry growth, those that start with techniques that are efficient only after the firm has grown considerably may be defeated in the evolutionary struggle by firms whose techniques are better suited to low levels of output’ (Nelson & Winter, 1982: 159). Building upon this insight, Denrell and March (2001) showed that practices that improve by learning-by-doing may be selected against because of their poor initial performance. In their simulation firms were endowed with one of two possible technologies. The first technology generated a fixed payoff. The second technology generated a low initial payoff but its payoff increases over time. Even if the long-term payoff of the second technology is higher than the fixed payoff of the first technology, implying that the second technology is ‘superior’, firms with the second technology were likely to fail before the potential of their technology was revealed. As a result, the proportion of firms with the second technology decreased over time. Selection is inefficient in this case because the proportion of firms with the second technology, which has a higher long-term payoff, is reduced over time. The inefficiency occurs because the fitness of the second technology changes systematically over time and the selection process is myopic in the sense that it reacts only to current payoffs and not to anticipated future payoffs.

Consider next frequency-dependent fitness. This refers to situations where the fitness of a gene depends on the relative abundance of that gene versus other genes in the environment. It is well known in evolutionary theory that when fitness is frequency dependent, a gene with lower potential fitness could become dominant (Wright, 1931; Maynard Smith, 1982). Evolutionary scholars in strategy and organization theory have similarly explored how the fitness of a particular practice may depend on the presence of other practices, creating a ‘rugged’ fitness landscape with multiple local optima in which selection is unlikely to identify the global optima (Arthur, 1989; Levinthal,

1997). Carroll & Harrison (1994) showed, in an evolutionary model, how frequency dependence allows inefficient organizational forms to survive. Following past empirical work, they assume that both the founding and mortality rates are frequency (density) dependent. Organizational forms only differ in their competitive effects: the negative effect that *one* organization of form  $i$  exerts on organizations of form  $j$ . An organizational form  $i$  is superior to an organizational form  $j$  if the competitive effect of  $i$  is larger than the competitive effect of  $j$ . Their simulations show that whenever an inferior form emerges earlier than a superior form it can become dominant. The intuition is that the inferior form exerts a larger *total* competitive pressure on the superior form, than vice versa, because there are more organizations of the inferior form (higher density). Simulations of their model show that when the inferior form enters first the density of the superior form initially increases but eventually declines to 0. The density of the superior form initially increases because, at this stage, the density of the inferior form is not yet very high. The decline occurs because when the inferior form becomes numerous it exerts a high total competitive pressure on the superior form. This scenario also represents a kind of ‘inefficient’ selection. The superior form is eliminated because there are more organizations of the inferior form and performance depends on the number (frequency) of organizations of the same form.

Variability is another reason why selection may not increase the proportion of a trait with the highest expected performance. It is well-known in evolutionary theory that a trait associated with the highest level of expected performance may be selected against if its performance is also highly variable (Cohen, 1966; Yoshimura & Clark, 1993; Cvijovic et al., 2015). Similarly, management theorists have argued that an organizational practice with the highest expected performance may decrease over time as a result of selection if this form also has the highest variance in performance. In particular, Levinthal & Posen (2007) and Levinthal & Marino (2015) have shown that adaptive learning processes may be selected against because they imply increased variability in performance as a result of the adjustments the firm will go through. Adaptive learning may lead to superior average performance in the long run but may lead to a higher variability in performance in the short run. Higher variability, in turn, increases the chances of elimination in the short-run, before the long-term advantages have been realized.

The mechanism we rely upon in this paper is different from past work because we do not assume that the performance effect of organizational characteristics is time-dependent or frequency-dependent. Nor do we assume that variability in performance is systematically related to expected performance. We assume instead that variability in performance is the same across all units in the population (the noise terms of all firms are drawn from the same distribution, independent of skill).

We show that even in this case there exist conditions under which selection will be inefficient in the sense that the proportion of agents with the superior trait does not necessarily increase over time. To the contrary, the proportion of agents with the superior trait ultimately *declines* over time.

It is important to note that our model does not imply that less ‘fit’ firms are more likely to survive. In each period selection removes firms with low performance levels, consistent with the idea of the ‘survival of the fittest’ if fitness is measured in terms of performance. Indeed, because poorly performing firms are replaced, average performance in the population increases over time, implying that the population becomes more ‘fit’ over time. The fact that the population becomes more ‘fit’ over time, however, is equally true for other mechanisms of inefficient selection, such as time-dependent fitness. Selection may reduce the proportion of the technology with highest long-term performance but average performance can nevertheless increase over time if the performance of the inferior technology also improves over time.

If selection leads to a monotonic performance increase over time in our model, why does it matter that the proportion of firms with high skill ( $u_i = 1$ ) decreases over time? For an observer interested only in average performance, the systematic decline in  $u_i$  over time may not matter. The decline in  $u_i$  over time would matter for an observer interested in identifying practices and skills that contribute to high performance. If  $u_i$  represents the capability of a firm, while  $\varepsilon_i$  represents situational influences beyond the control of management, such an observer would be interested in learning from firms with high capabilities ( $u_i = 1$ ). Our results imply that such an observer should not imitate firms that have been through many rounds of selection. In Section 8 we discuss in more detail when our results do and do not matter.

#### 4. FORMAL ANALYSIS I: SELECTION AND NO REPLACEMENT

To analyze when and why selection can reduce the proportion of the type with the highest value of  $u_i$  we first focus on a simple setting when there is selection but no replacement. That is, we analyze the effect of repeated selection on a cohort of agents. In Section 6.1 we show that the basic result continues to hold if there is also replacement.

**4.1. Model.** Consider a population of infinitely many agents (an agent can be an individual, a firm etc.). Assuming that a population is made of infinitely many agents is a standard assumption in many evolutionary models that allows for mathematical tractability. Our illustrative graphs in Section 2 showed that we get similar results in a population of 50 agents.

Agents are either high skill agents ( $u_i = 1$ ) or low skill agents ( $u_i = 0$ ). The level of skill of an agent remains the same during its lifetime. Initially, the proportion of high skill agents is

0.50. We use the label ‘skill’ to denote a trait that contributes to the performance of agent  $i$ . The performance of agent  $i$  in period  $t$ ,  $p_{i,t}$ , equals her skill,  $u_i$  plus a noise term,  $\varepsilon_i$ :  $p_{i,t} = u_i + \varepsilon_{i,t}$ . The noise term represents an aspect of performance beyond the control of the agent. We assume that the noise term is drawn from a density  $f_\varepsilon$  with positive variance. We assume that the support of  $f_\varepsilon$  is of the form  $(a, +\infty)$  with  $a \in \{-\infty, \mathbb{R}\}$ .<sup>1</sup> Note that while we use the label ‘noise term’ for  $\varepsilon_i$ , we do not assume that the expected value of  $f_\varepsilon$  has to be equal to 0.

Selection works as follows. In each period  $t$ , the  $w$  percent agents with the lowest performance levels (lowest values of  $p_{i,t}$ ) are removed from the population.

We denote by  $\pi_t$  the proportion of high skill agents at the end of period  $t$ . We denote by  $\pi_0$  the initial proportion of high skill agent ( $\pi_0 = 0.50$ ). We are interested in how  $\pi_t$  evolves with  $t$ . Does it increase with  $t$  implying that selection increases the proportion of agents with high skill?

**4.2. Independent Noise Terms.** Suppose the noise terms are independent across agents and periods. That is,  $p_{i,t} = u_i + \varepsilon_{i,t}$ , and  $\varepsilon_{i,t}$  are *iid* draws from the density  $f_\varepsilon$  for all  $t$  and  $i$ . We denote the density of the performance distribution of high skill agents by  $f_1$  and the density of the performance distribution of low skill agents by  $f_0$ . Because the noise terms are redrawn in every period, the distribution of performance conditional on skill remains constant. Hence, we have:

$$f_0(p_{i,t+1}) = f_0(p_{i,t}),$$

and

$$f_1(p_{i,t+1}) = f_1(p_{i,t}).$$

It follows that, in every period, high skill agents (whose performances equal  $p_{i,t} = 1 + \varepsilon_{i,t}$ ) are likely to have a higher performance than low skill agents (whose performances equal  $p_{i,t} = \varepsilon_{i,t}$ ). As a result, the proportion of high skill agents increases over time. Theorem 1 demonstrates that this holds for any (continuous) distribution of the noise term:

**Theorem 1.** *Suppose  $p_{i,t} = u_i + \varepsilon_{i,t}$ , where, for all  $t$  and  $i$ ,  $\varepsilon_{i,t}$  are iid draws from the continuous density  $f_\varepsilon$ . Whenever  $w \in (0, 1)$  expected skill increases over time:  $\pi_{t+1} > \pi_t$  for all  $t \geq 1$ .*

*Proof.* See Appendix A. □

**4.3. Constant Noise Terms.** Suppose now that the noise terms remain the same in all periods. The noise term in the first period  $\varepsilon_{i,1}$  is drawn from density  $f_\varepsilon$ . The noise terms in periods 2, 3, 4, ...

<sup>1</sup>Ancillary analyses and computer simulations show that most of our results still apply when the support of  $f_\varepsilon$  has an upper bound, but assuming that there is an upper bound leads to much more complicated proofs with many special cases that makes it difficult to understand the mechanism at the core of our analysis.

are identical to the noise term drawn in period 1. The performance of agent  $i$  thus remains the same in all periods: For all  $t$ ,  $p_{i,t} = u_i + \varepsilon_{i,1}$ . This setup represents a situation of extreme dependency across periods. Such dependency can occur when initial ‘luck’ has long-term effects.

An important implication of this specification is that the distribution of performance conditional on skill and survival systematically changes over time. The reason is that selection in prior periods eliminates agents with low values of  $p_{i,t}$  and thus indirectly eliminates agents with low values of  $\varepsilon_{i,1}$ . Whether selection increases average skill or not then depends on the nature of the distribution of the noise term ( $f_\varepsilon$ ).

The following theorem shows that whether selection leads to a monotonic increase in the proportion of high skill agents or not depends on whether the hazard function of the noise distribution is an increasing function or not. The hazard function  $h$  of a distribution  $f$  is defined as the ratio of the density over 1 minus the cumulative density function  $F$ :  $h(x) = f(x)/(1 - F(x))$ .

**Theorem 2.** *Suppose  $p_{i,t} = u_i + \varepsilon_{i,1}$  and  $\varepsilon_{i,1}$  is drawn from density  $f_\varepsilon$ . Let  $h_\varepsilon$  denote the hazard function of distribution  $f_\varepsilon$ . In this case,*

- i) The proportion of high skill agents increases as a result of selection during the first period:  $\pi_1 > \pi_0 = 0.5$*
- ii) The evolution of the proportion of high skill agents after the first period ( $t > 1$ ) depends on the shape of the hazard function of the noise distribution:*
  - a) If  $h_\varepsilon$  is an increasing function, then  $\pi_t$  increases with  $t$ .*
  - b) If  $h_\varepsilon$  is a decreasing function, then  $\pi_t$  decreases with  $t$  for  $t$  large enough,*
  - c) If there exists  $c^*$  such that  $h_\varepsilon(c - 1) < h_\varepsilon(c)$  for all  $c < c^*$  and  $h_\varepsilon(c - 1) > h_\varepsilon(c)$  for all  $c > c^*$ ,  $\pi_t$  decreases with  $t$  for  $t$  large enough.*
  - d) If there exists  $c^*$  such that  $h_\varepsilon(c - 1) = h_\varepsilon(c)$  for all  $c > c^*$ ,  $\pi_t$  remains constant for  $t$  large enough.*

*Proof.* See Appendix B. □

Table 4.3 lists several distributions and whether their hazard functions are increasing or decreasing (based on Bagnoli & Bergstrom (2005) and Glaser (1980)). Table 4.3 shows that selection always increases average skill whenever the noise term is drawn from a normal distribution, the logistic distribution, the extreme value distribution, and the Weibull distribution (density  $k\epsilon_i^{k-1}e^{-\epsilon_i^k}$ ) with parameter  $k > 1$ . Selection eventually decreases average skill, however, for several distributions with ‘fatter’ tails than the normal distribution, including the t-distribution, the Cauchy distribution, the

Log-normal distribution, the Inverse Gaussian, the Weibull distribution with parameter  $k < 1$ , and the Pareto distribution. There exists distributions, such as the Laplace distribution, which has fatter tails than the normal distribution but which nevertheless does not have a decreasing hazard function.

TABLE 1. Shape of the hazard functions of a set of distributions.

Distribution	Shape of the Hazard Function
Uniform	Increasing
Normal	Increasing
Logistic	Increasing
Poisson	Increasing
Extreme Value	Increasing
Exponential	Constant
Laplace	Initially increasing, eventually constant
Cauchy	Initially increasing, eventually decreasing
Log-normal	Initially increasing, eventually decreasing
Inverse Gaussian	Initially increasing, eventually decreasing
Weibull	Increasing when $k > 1$ , decreasing when $k < 1$
Pareto	Decreasing

**4.4. Long-Tailed Distributions.** Theorem 2 shows that the proportion of high skill agents will eventually decline over time when the noise term is drawn at time of entry from a distribution with a hazard function which is (eventually) decreasing. This raises the question of which distributions have a declining hazard function. Can these be characterized in some intuitive way? In this section we show that a class of distributions, called ‘long-tailed’ distributions, have exactly the properties we seek.

Formally, a random variable  $X$  has a long-tailed distribution if  $(1 - F_x(c + y))/(1 - F_x(c)) \rightarrow 1$  as  $c \rightarrow \infty$  for all  $y > 0$  (Foss et al., 2013). Here  $F_x()$  denotes the cumulative distribution function of  $X$ . The property of having a long tail thus corresponds to the fact that if  $X$  is larger than some very large constant  $c$  (which occurs with probability  $1 - F_x(c)$ ) then  $X$  is also likely to be larger than  $c + y$  (which occurs with probability  $1 - F_x(c + y)$ ). Examples of long-tailed distributions include the t-distribution with one degree of freedom (i.e, the Cauchy distribution) but also the Pareto distribution and the Log-normal distribution.

The property of being a long-tailed distributions turns out to be the property that precisely characterizes the set of distributions for which survival during many periods eventually becomes uninformative about skill when the noise terms are constant. Formally, let  $p_t^*$  be the maximum

level of performance among the agents that were eliminated in period  $t$ . All survivors during period  $t$  have a performance above  $p_t^*$ . When the noise terms remain constant the proportion of high skill agents among the survivors of  $t$  periods is  $P(u_i = 1 \mid p_i > p_t^*)$ , the proportion of high skill agents among agents that have a performance above  $p_t^*$ . The property of being long-tailed precisely captures the set of noise term distributions for which  $P(u_i = 1 \mid p_i > p_t^*)$  becomes uninformative about skill as  $p_t^*$  becomes large. Formally,

**Theorem 3.** *Let  $p_i = u_i + \varepsilon_i$  where  $u_i = 1$  with probability 0.5 and  $u_i = 0$  otherwise. Then,*

- i)  $\lim_{c \rightarrow \infty} P(u_i = 1 \mid p_i > c) = 0.5$  if and only if  $f_\varepsilon$  is the density of a long-tailed distribution.*
- ii)  $\lim_{t \rightarrow \infty} P(u_i = 1 \mid p_i > p_t^*) = 0.5$  if and only if  $f_\varepsilon$  is the density of a long-tailed distribution.*

*Proof.* See Appendix C. □

Theorem 3 implies that the proportion of high skill agents will converge to 50%, the initial proportion, when the noise term is drawn from a ‘long-tailed’ distribution.

What about the condition regarding the hazard function? Theorem 2 states that average skill eventually declines if the hazard function of the noise term distribution is eventually declining. It turns out that the property of having a long tail also implies that the hazard function is (eventually) declining. Formally, if a distribution is long-tailed then its hazard function  $h(x)$  will eventually go to 0:  $h(x) \rightarrow 0$  as  $x \rightarrow \infty$  (Nair et al., 2013).

How does the property of being long-tailed relate to the more widely known concept of ‘fat-tailed’ distributions? A distribution is ‘fat-tailed’ if the upper tails behaves as a power law, i.e., if  $P(x > c) \approx c^{-x}$  as  $c \rightarrow \infty$  (Foss et al., 2013). Fat-tailed distributions belong to the class of ‘heavy’ tailed distributions (all fat-tailed distributions are heavy-tailed but there exists heavy-tailed distributions which are not fat-tailed). Informally, a distribution is heavy-tailed if its tail is heavier than the tail of an exponential distribution.<sup>2</sup> Long-tailed distributions are a sub-class of ‘heavy-tailed’ distributions (Nair et al., 2013). Hence, all long-tailed distributions are heavy tailed. There exist long-tailed distributions, however, which are heavy-tailed but not fat-tailed (an example is the Log-normal distribution).

## 5. INTUITION

Why does selection increase the proportion of high skill agents if the noise terms are redrawn but can decrease the proportion of high skill agents if the noise terms are constant? And why does

---

<sup>2</sup>Formally, a distribution is heavy-tailed if the moment generating function is infinite (Foss et al., 2013).

the effect of selection depend on the hazard function of the noise term? The basic result can be explained in two different ways.

**5.1. Intuition 1: The Diagnosticity of Survival Decreases with Time.** The first type of explanation focuses on the skill levels of the survivors. After several periods of selection only a small fraction of the initial population remains. These survivors have had high performance during several periods. Is such high performance a reliable indication of high skill? This is not generally the case. It is well-known in statistics that a higher outcome may not indicate a higher expected value (Karlin & Rubin, 1956). For some ‘heavy-tailed’ noise term distributions a very high outcome may indicate a lower expected value than a moderately high outcome does (Weibull et al., 2007; Denrell & Liu, 2012). The reason is that extreme outcomes depend relatively more than moderately high outcomes on luck than skill. Survival during many period can, for similar reasons, be an unreliable indicator of high skill.

To explain this, consider the agents that have survived during all of the first  $t$  periods. What is the proportion of high skill agents among these survivors? Consider first the case when noise terms are constant. Performance in consecutive periods does not change (because the noise term remain constant). Moreover, because selection removes the agents with the lowest performances, the threshold for survival,  $p_t^*$ , defined as the maximum level of performance among the agents that failed in period  $t$ , increases over time (See Lemma 4 in the Appendix). This implies that if an agent had a performance above the threshold in period  $t$  ( $p_{i,t} > p_t^*$ ), her performance was also above the threshold in any previous period ( $p_{i,t-1} > p_{t-1}^*$ ). The probability of surviving  $t$  periods for an agent with skill  $u_i = k$  is thus simply the probability that the performance drawn in the first period is above the threshold in period  $t$ :  $P(p_{i,t} > p_t^* \mid u_i = k)$ . When the noise terms are constant, the proportion of high skill agents among the survivors is thus

$$\pi_t = P(u_i = 1 \mid p_{i,t} > p_t^*) = \frac{P(p_{i,t} > p_t^* \mid u_i = 1)P(u_i = 1)}{P(p_{i,t} > p_t^* \mid u_i = 1)P(u_i = 1) + P(p_{i,t} > p_t^* \mid u_i = 0)P(u_i = 0)}$$

When the noise terms are constant the only thing we know about the agents that have survived for  $t$  periods is that the performance they drew at the start was above  $p_t^*$ , the threshold for survival in period  $t$ . When  $t$  is large, the value of  $p_t^*$  will be high. Hence, we know that the performance levels of all the survivors are very high. But the fact that they all have a high level of performance does not imply that they are all high skill agents. It is not generally true that the proportion high skill agents is higher among agents with higher levels of performance.

This is illustrated in Figure 4. The upper quadrants show that the threshold for survival increases over time (when  $w = 0.1$ ) both when the noise term is drawn from a normal distribution and from



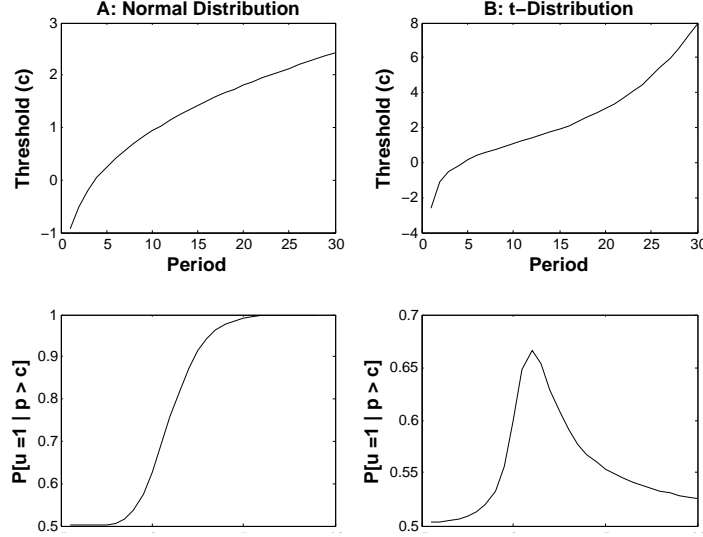


FIGURE 4. Upper panels: Survival threshold as a function of time. Lower quadrants: Proportion of high skill agents among those with a performance above the threshold. Left panels: Noise term is drawn from a Normal distribution with mean 0 and variance 1. Right quadrants: Noise terms are drawn from a t-distribution with mean 0 and 1 degree of freedom.

a t-distribution. The lower quadrant shows how the proportion of high skill agents that have a performance above a threshold  $P(u_i = 1 \mid p_{i,t} > c)$  varies with the threshold,  $c$ . When the noise term is drawn from a normal distribution  $P(u_i = 1 \mid p_{i,t} > c)$  increases with  $c$ . When the noise term is drawn from a t-distribution  $P(u_i = 1 \mid p_{i,t} > c)$  eventually declines with  $c$  and converges to 0.5 as  $c \rightarrow \infty$ . As Theorem 3 shows,  $\lim_{c \rightarrow \infty} P(u_i = 1 \mid p_{i,t} > c) = 0.5$  if and only if  $f_\varepsilon$  is the density of a long-tailed distribution.

The situation differs when the noise terms are independently drawn in each period. In this case, survival during many periods does not become uninformative. To explain why, note that in this case performance changes every period because the noise terms are redrawn in every period. Moreover, conditional upon skill, performance levels in different periods are independent. The probability of survival during  $t$  periods for an agent with skill  $u_i = k$  is thus the probability of surviving period 1,  $P(p_{i,1} > p_1^* \mid u_i = k)$ , multiplied by the probability of surviving period 2,  $P(p_{i,2} > p_2^* \mid u_i = k)$ , etc. The proportion of high skill agents among the survivors of  $t$  periods can thus be written as

$$\pi_t = \frac{P(p_{i,1} > p_1^*, \dots, p_{i,t} > p_t^* \mid u_i = 1)P(u_i = 1)}{P(p_{i,1} > p_1^*, \dots, p_{i,t} > p_t^* \mid u_i = 1)P(u_i = 1) + P(p_{i,1} > p_1^*, \dots, p_{i,t} > p_t^* \mid u_i = 0)P(u_i = 0)},$$

Noting that  $P(u_i = 1) = P(u_i = 0) = 0.5$  and that performance levels in successive periods are independent from each other conditional on skill, we can write:

$$\pi_t = \frac{\prod_{j=1}^t P(p_{i,j} > p_j^* \mid u_i = 1)}{\prod_{j=1}^t P(p_{i,j} > p_j^* \mid u_i = 1) + \prod_{j=1}^t P(p_{i,j} > p_j^* \mid u_i = 0)}.$$

In contrast to the case of constant noise terms, this expression does not only depend on the probability that the final performance is above the final threshold,  $P(p_{i,t} > p_t^* \mid u_i = k)$ , but also on what happened in all prior periods. Moreover, because the performance in period  $t$  is independent from performance in earlier periods, conditional on skill, more information is available compared to the case when the noise terms are constant. Stated differently, when the noise terms are independent an agent has to pass  $t$  independent tests to survive during  $t$  periods. While the test in the final period may be relatively uninformative about skill, because  $p_t^*$  is very high in that period, the early tests, when  $p_t^*$  is moderately high, remain informative.

**5.2. Intuition 2: Failure Rates Conditional on Skill Change over Time.** While the first explanation that focused on the diagnosticity of survival provides an intuition for why the proportion of high skill agents may eventually decrease when noise terms are constant and drawn from a long-tailed distribution, it does not explain why the hazard function of the noise term matters. We now provide an intuition for this.

In every period,  $w$  percent of all agents ‘fail’ and are removed from the population. The change in the proportion of high skill agents depends on the proportion of high skill and low skill agents among the failures. Suppose, for example, that almost all of the agents that fail have low skill. In this case, it is intuitively clear that the proportion high skill agents among the survivors will increase. More generally, if the probability of failure is higher for low skill agents, the proportion of low skill agents decreases and hence average skill increases (see Lemma 1 in Appendix A).

To understand when selection increases the proportion of low skill agents, we thus need to understand when low skill agents are, in every period, less likely to fail than high skill agents. An agent fails whenever its performance is among the  $w$  percent lowest performances. If  $p_{t+1}^*$  is defined as the maximum performance among the failures in period  $t + 1$ , we can say that an agent fails in period  $t + 1$  whenever its performance in period  $t + 1$  is at or below  $p_{t+1}^*$ .

The difference between high and low skill agent can then be stated as follows. 1) A low skill agent  $j$  (with skill equal to  $u_j = 0$ ) fails in period  $t + 1$  whenever  $\varepsilon_{j,t+1} \leq p_{t+1}^*$ . The probability of failure in period  $t + 1$  for a low skill agent is thus the probability that  $\varepsilon_{j,t+1}$  is at or below  $p_{t+1}^*$ . 2) A high skill agent  $i$  (with skill equal to  $u_i = 1$ ) fails in period  $t + 1$  whenever  $1 + \varepsilon_{i,t+1} \leq p_{t+1}^*$ .

Formulated differently: a high skill agent  $i$  fails when  $\varepsilon_{i,t+1}$  is at or below  $p_{t+1}^* - 1$ . The probability of failure in period  $t + 1$  for a high skill agent is thus the probability that  $\varepsilon_{i,t+1}$  is at or below  $p_{t+1}^* - 1$ . This formulation holds the key to understanding the difference between independent and constant noise terms.

Consider the case with independent noise terms. The noise terms are, for all agents and all periods, drawn from the same distribution. So the noise term for a high skill agent  $i$ ,  $\varepsilon_{i,t+1}$ , is drawn, in all periods  $t + 1$ , from a distribution identical to that of  $\varepsilon_{j,t+1}$ , the noise term for a low skill agent  $j$ . Because  $\varepsilon_{i,t+1}$  and  $\varepsilon_{j,t+1}$  are independently drawn from the same distribution it is clear that the probability the low skill agent fails is higher than the probability that the high skill agent fails:  $P(\varepsilon_{j,t+1} \leq p_{t+1}^*) > P(\varepsilon_{i,t+1} \leq p_{t+1}^* - 1)$ . Because this holds for any period  $t + 1$ , the proportion of high skill agents always increases (as long as the proportion of high skill agents is below 1).

Consider now the case with constant noise terms. Let us first focus on the probability that a low skill agent fails in period 2. A low skill agent  $j$  fails in period 2 (and not in period 1) if its performance is sufficiently high in period 1 and its performance in period 2 is at or below the threshold for survival. Formally, a low skill agent  $j$  fails in period 2 if  $p_{j,1} = \varepsilon_{j,1} > p_1^*$  and  $p_{j,2} = \varepsilon_{j,1} \leq p_2^*$ . What matters for failure in period 2 is thus the distribution of performance in period 2 *conditional on survival during period 1*.

Suppose now that  $p_2^* = p_1^* + y$ , where  $y > 0$ . Agents with a performance above  $p_1^*$  and at or below  $p_1^* + y$ , for some  $y > 0$ , fail in period 2. The probability that a low skill agent  $j$ , who survived period 1, fails in period 2 falling is thus

$$P(p_1^* + y \geq p_{j,1} > p_1^* \mid p_{j,1} > p_1^*, u_j = 0) = \frac{P(p_1^* + y \geq p_{j,1} > p_1^* \mid u_j = 0)}{P(p_{j,1} > p_1^*, u_j = 0)}.$$

Because  $p_{j,1} = \varepsilon_{j,1}$  when  $u_j = 0$  this conditional probability equals

$$(1) \quad P(p_1^* + y \geq p_{j,1} > p_1^* \mid p_{j,1} > p_1^*, u_j = 0) = \frac{P(p_1^* + y \geq \varepsilon_{j,1} > p_1^*)}{P(\varepsilon_{j,1} > p_1^*)}.$$

Imagine now that  $w$  is small. Only a small fraction of all agents are eliminated in each period. Thus only agents with a performance just above to  $p_1^*$  (the threshold in period 1) are eliminated during period 2. That is,  $y$  is close to 0. Formally, let  $y \rightarrow 0$  in equation 1. As  $y \rightarrow 0$ , the right hand side converges to  $h_\varepsilon(p_1^*) = f_\varepsilon(p_1^*)/P(\varepsilon_{j,1} > p_1^*)$ , the hazard function of the noise distribution at  $p_1^*$  (Ross (2000), p. 220).

Now, let us focus on the probability that a high skill agent fails in period 2. A high skill agent  $i$  fails in period 2 if  $p_{i,1} = 1 + \varepsilon_{i,1} > p_1^*$  and  $p_{i,2} = 1 + \varepsilon_{i,1} \leq p_2^*$ . The probability that a high skill

agent fails in period 2 thus depends on the distribution of performance conditional upon survival during period 1:  $f_{p_{i,1}|p_{i,1}>p_1^*}(p_{i,1} | p_{i,1} > p_1^*)$ . Again, suppose that  $p_2^* = p_1^* + y$ . The probability that a high skill agent  $i$ , who survived period 1, fails in period 2 is

$$P(p_1^* + y \geq p_{i,1} > p_1^* | p_{i,1} > p_1^*, u_i = 1) = \frac{P(p_1^* + y \geq p_{i,1} > p_1^* | u_i = 1)}{P(p_{i,1} > p_1^*, u_i = 1)}.$$

Because  $p_{i,1} = 1 + \varepsilon_{i,1}$ , this conditional probability equals

$$P(p_1^* + y \geq p_{i,1} > p_1^* | p_{i,1} > p_1^*, u_i = 1) = \frac{P(p_1^* + y - 1 \geq \varepsilon_{i,1} > p_1^* - 1)}{P(\varepsilon_{i,1} > p_1^* - 1)}.$$

As  $y \rightarrow 0$  this expression converges to  $h_\varepsilon(p_1^* - 1) = f_\varepsilon(p_1^* - 1)/P(\varepsilon_{i,1} > p_1^* - 1)$ , the hazard function of the noise distribution at  $p_1^* - 1$ .

In summary, whether low skill or high skill agents are more likely to fail during period 2 depends on the hazard function of the noise term distribution. Low skill agents are more likely to fail than high skill agents are if the hazard function at  $p_1^*$ ,  $h_\varepsilon(p_1^*)$ , is larger than the hazard function at  $p_1^* - 1$ ,  $h_\varepsilon(p_1^* - 1)$ . It follows that low skill agents are more likely to fail than high skill agents whenever  $h_\varepsilon(x) > h_\varepsilon(x - 1)$ , i.e., whenever the hazard function of the noise distribution is an increasing function.

This result can be restated in terms of the hazard functions for the performance distributions of low and high skill agents. The hazard function of the performance distribution for a low skill agent,  $h_{p_{j,t},u=0}(x)$ , is equal to the hazard function for the noise distribution,  $h_\varepsilon(x)$  since  $p_{j,t} = \varepsilon_{j,1}$  for low skill agents. The hazard function of the performance distribution for a high skill agent,  $h_{p_{i,t},u=1}(x)$ , is equal to the hazard function for the noise distribution evaluated at  $x - 1$ ,  $h_\varepsilon(x - 1)$ , since the equation  $\varepsilon_{j,1} = p_{i,t} - 1$  holds for high skill agents. It follows that the inequality  $h_\varepsilon(x) > h_\varepsilon(x - 1)$  for all  $x$  can be restated as  $h_{p_{j,t},u=0}(z) > h_{p_{i,t},u=1}(z)$  for all  $z$ , i.e., the hazard function for low skill agents always lies above the hazard rate function for high skill agents.

As Figure 5A shows, the hazard function is an increasing function when the noise term is drawn from a normal distribution. Moreover, the hazard rate function for low skill agents always lies above the hazard rate function for high skill agents. As a result, a low skill agent is, in every period, more likely to fail than a high skill agent. To demonstrate this, Figure 6 plots how the probability of failure varies over time for high and low skill agents (for the case when there is only selection and no replacement). Specifically, Figure 6 plots the probability of failure conditional on survival,  $P(p_{i,t} \leq p_t^* | \forall j < t : p_{i,j} > p_j^*, u_i = k)$  for high and low skill agents. The probability of failure is always higher for low skill agents when the noise terms are drawn from a normal distribution (Figure 6A). When the noise term is drawn from a t-distribution, however, the hazard function for

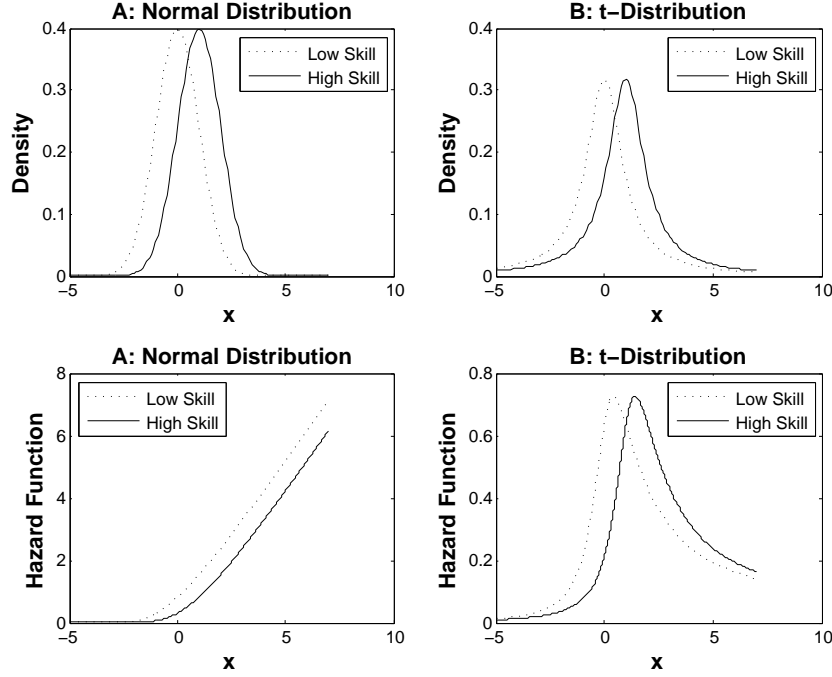


FIGURE 5. Upper panels: Density of performance for low and high skill agents when the noise distribution is drawn from A) a normal distribution and B) a t-distribution. Lower panels: hazard functions for high and low skill agents.

the high skill agents eventually becomes higher than the hazard function for the low skill agents (Figure 5B). The implication of such crossing hazard function is that the probability of failure will eventually become higher for high skill agents than for low skill agents (Figure 6B).

Why do high skill agents eventually become more likely to fail than low skill agents when the noise term is drawn from a t-distribution and remains constant? The reason is that, when the noise term is drawn from such a ‘long-tailed’ distribution, low skill agents who have survived tend to have very high performances. Of course, low skill agents initially tend to have lower performances than high skill agents. As a result, low skill agents are initially more likely to fail. Once these poor performers are removed from the population, however, the remaining low skill agents, who have survived the early periods, tend to have very high performance. Because there has been less selection operating on high skill agents - their high skill buffers them from selection - there may in fact be more high skill agents with a medium level of performance. Because the performance threshold increases over time, it becomes more likely that high skill agents have a performance level just above the threshold. Eventually, when many high skill agents have been removed, the

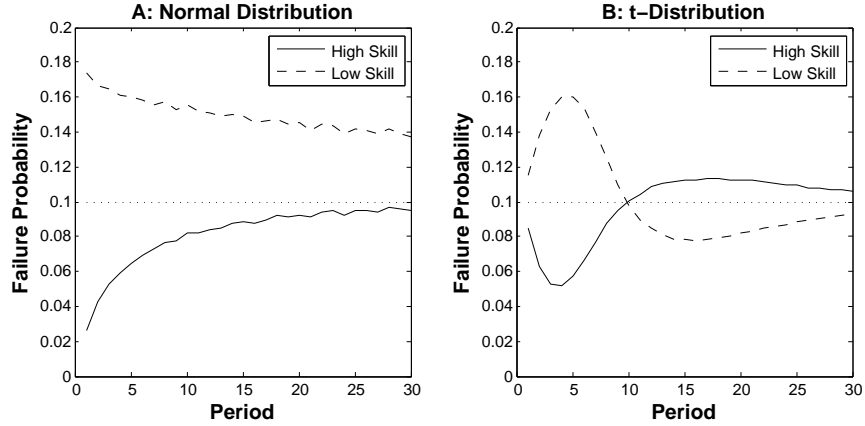


FIGURE 6. How the failure probability (when there is no replacement) varies over time for high and low skill agents when the noise terms are constant and drawn from A) a normal distribution and B) a t-distribution. Minor oscillations in the graphs are due to numerical imprecision in the computations.

surviving high and low skill agents have almost identical performance distributions and hence are equally likely to fail.

**5.3. Relation to Prior Work on Luck and Performance.** The mechanism that drives our main result is related to the mechanisms discussed in prior work on the role of luck in explaining performance patterns. Specifically, it is known from prior work that an extremely high level of performance may not be diagnostic of high skill when the noise term is drawn from a fat-tailed distribution such as the t-distribution (Weibull et al., 2007; Denrell & Liu, 2012). In this sense, our results follow from prior work. Nevertheless, our results are not a direct consequence of past work. For one thing, our analytical results (Theorems 1 and 2) enable us to go beyond prior work in characterizing precisely for which noise distributions average skill may decline over time. Note also that our results in this paper focus on the expected skill conditional performance *above* a given level (the performance threshold) while past work has focused on expected skill conditional on performance *equal to* a given level; a subtle difference but important formally. More important, this paper shows that past work about the conditions under which extreme performance may not be diagnostic of high skill is only relevant in settings where the noise terms are dependent (noise terms do not change or are added up, like in a random walk). Whenever the noise terms are redrawn,

survival during many periods, even if it would require extreme high levels of performance, is still diagnostic of high skill.

## 6. ALTERNATIVE ASSUMPTIONS ABOUT SELECTION AND REPLACEMENT

The model on which our formal analysis has focused is admittedly a toy model that hardly maps onto any naturally occurring settings. Do our results also hold in more realistic settings? Here we explore this issue by simulating models with different assumptions about how selection and replacement operates.

**6.1. Replacement.** Our main result about the emergence of a non-monotonic pattern in the proportion of high-skill agents also holds when the agents who exit the population are replaced by new agents. More specifically, consider the setting of Section 4.3 where the noise terms are constant. Here, we assume that everything remains the same, except that each eliminated agent  $i$  is replaced by a new agent  $j$  with performance  $p_j = u_i + \varepsilon_j$  where  $\varepsilon_j$  is independently drawn from density  $f_\varepsilon(\varepsilon)$ . The probability that a new agent has high skill ( $u_j = 1$ ) is 50%. We have:

**Theorem 4.** *Theorem 2 holds in this setting as well.*

*Proof.* See Appendix D. □

**6.2. Replacement from a changing skill pool.** We have assumed that the probability that a new agent has high skill is 50% in all periods. Entrants may learn from survivors, however. If so, the probability that an entrant has high skill may increase over time. To model this, suppose the probability that a new agent entering in period  $t$  has high skill equals  $q_t = r\pi_t + (1 - r)0.5$ . The probability that a new agent entering in period  $t$  has high skill is thus a weighted average between the proportion of agents with high skill among the survivors ( $\pi_t$ ) and 50%. The weight on  $\pi_t$  is  $r \in [0, 1]$ . It can be interpreted as the probability that an entrant is able to copy a survivor. Simulations show that the basic result holds even if  $r > 0$  as long as  $r$  is not too close to 1. Figure 7 shows how the proportion of high skill agents changes over time in a simulation with  $n = 50$  firms, the noise terms are drawn from a t-distribution with 1 degree of freedom,  $w = 0.1$ , and A)  $r = 0.7$ , B)  $r = 0.9$  and C)  $r = 1$ . The proportion of high skill agents eventually declines unless  $r = 1$ .

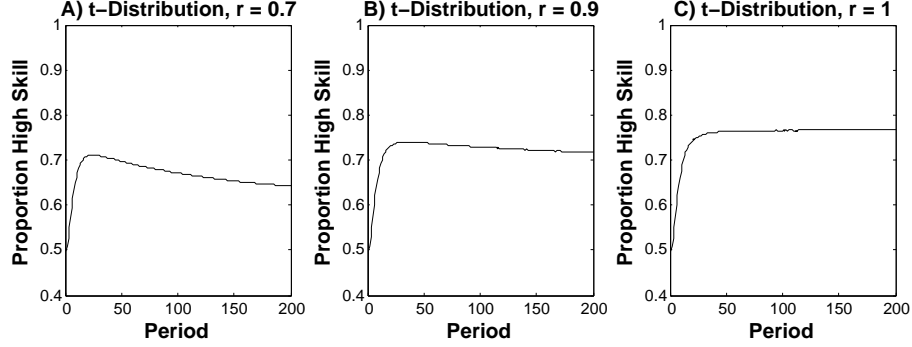


FIGURE 7. How the proportion of high skill agents changes over time in a simulation when there is replacement from a changing skill pool, the noise terms are drawn from a t-distribution with 1 degree of freedom,  $w = 0.1$ , and A)  $r = 0.7$ , B)  $r = 0.9$  and C)  $r = 1$  (based on 10,000 simulations, each with 50 agents).

**6.3. Probabilistic selection.** The models examined so far assumed that all agents with the lowest  $w$  percent performances were eliminated and replaced in each period. It may be more realistic to assume that selection operates in a probabilistic fashion: agents with relatively low performance are more likely than agents with relatively good performance to fail and be replaced. To model this, let  $p_t^w$  be the level of performance such that  $w$  percent of all agents have a lower level of performance in period  $t$ . Previously we assumed that all agents with a performance at or below  $p_t^w$  were eliminated and all agents with a performance above  $p_t^w$  survived. To model probabilistic selection we assume instead that the probability that agent  $i$  survives period  $t$  is  $1/(1 + \exp(-s(p_{i,t} - p_t^w)))$ . Here  $s \geq 0$  is a parameter that regulates the extent to which selection is sensitive to relative performance. A larger value of  $s$  implies that the probability of survival is more sensitive to relative performance ( $p_{i,t} - p_t^w$ ). When  $s \rightarrow \infty$  all agents with a performance above  $p_t^w$  survive and all agents with a performance below that threshold are replaced. When  $s = 0$ , the probability of survival equals 0.5 for all agents and is thus unrelated to relative performance.

Simulations show that the basic result hold even for probabilistic selection unless the value of  $s$  is low. To illustrate this, suppose  $w = 0.1$ . Here,  $p_t^w$  is equal to the tenth percentile of the performance distribution. Figure 8A shows the survival probability, in the first period, for different percentiles of the performance distribution when  $s = 1$  and  $s = 0.1$ . The survival probability is



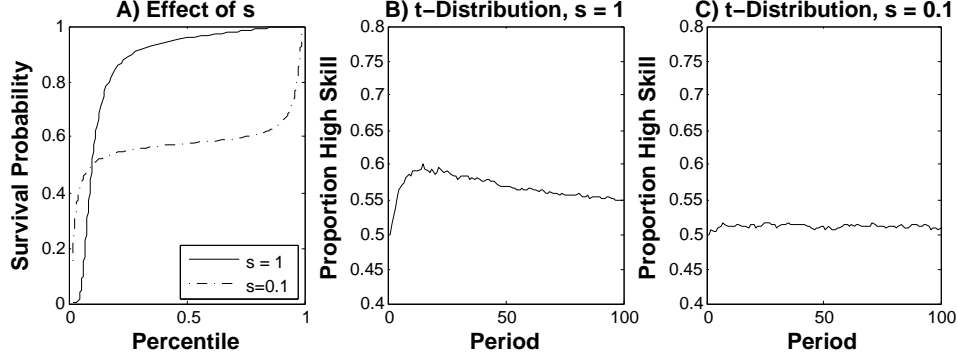


FIGURE 8. A) The impact of the value of  $s$  on the probability of survival. B) and C) How the proportion of high skill agents changes over time when the noise term is drawn from a t-distribution with B)  $s = 1$  and C)  $s = 0.1$  (based on 10,000 simulations, each with 50 agents).

0.5 for a performance level equal to the tenth percentile (i.e.,  $p_{i,1} = p_1^{0.1}$ ). If  $s = 1$  the survival probability quickly increases towards one for higher performance levels while if  $s = 0.1$  the survival probability remains at a moderate level (around 55%) unless performance is very high. Panels B and C in Figure 8 show how the proportion of high skill agents changes over time when  $s = 1$  and  $s = 0.1$ . In each simulation there are  $n = 50$  agents, high and low skill are initially equally likely, the noise terms are constant during the life-time of an agent, and each eliminated agent is replaced with a new agent. The proportion of high skill agents initially increases but eventually declines when  $s = 1$ . When  $s = 0.1$ , and even high performing agents may fail to survive, the proportion initially increases but then reaches a plateau.

**6.4. Fixed Selection Threshold.** The proportion of high skill agents only decreases over time when the performance required for survival increases over time. The assumption that selection operates on relative performance - the agents with the  $w$  percent lowest levels of performances fail and are replaced - ensures that the performance required for survival does increase over time. Suppose, alternatively, that all agents with a performance above a fixed cutoff survive. In such a model the proportion of high skill agents monotonically increases over time (if there is no replacement) or increases over time until it reaches an equilibrium (if there is replacement) and never decreases.

The implication is that the proportion of high skill agents can only decrease if survival depends on relative rather than absolute performance.

**6.5. Size Changes.** Our model focused on the elimination of poorly performing units and assumed that all agents were of equal size. In contrast, several evolutionary models in management have focused on growth. For example, Nelson and Winter (1982) explored the consequences of the assumption that profitable firms grow and unprofitable firms contract. Changes in traits among survivors may be relatively uninteresting in such a model. What matters are size-weighted traits: whether large firms are more likely to use ‘efficient’ technologies.

Do our results hold also in a model where relative performance determines growth? To explore this, we simulated a simple version of a model in which relatively high performance leads to growth in size while relatively low performance leads to contraction in size. The model is specified as follows. There are  $n = 50$  firms. The performance of firm  $i$  is  $p_i = u_i + \varepsilon_i$ . There are only two levels of  $u_i$ : high ( $u_i = 1$ ) and low ( $u_i = 0$ ), each equally likely. The level of  $u_i$  remains the same during the life-time of a firm. So does the noise term,  $\varepsilon_i$ . There is no selection in this model: we ignore selection to focus on size changes. All firms survive all periods and there are no entrants. Firms do change in size, however. At the start, each firm has size  $s_{i,1} = 1$ . Firm size changes, in each period, as follows. Let  $z_t$  be the market share weighted performance in period  $t$ :  $z_t = \sum_{i=1}^n m_{i,t} p_i$  where  $m_{i,t} = s_{i,t} / \sum_{i=1}^n s_{i,t}$  is the market share of firm  $i$  at the start of period  $t$ . Every firm with performance above  $z_t$  increases in size by 10%: if  $p_i > z_t$  then  $s_{i,t+1} = 1.1s_{i,t}$ . Every firm with performance below  $z_t$  contracts in size by 10%: if  $p_i < z_t$  then  $s_{i,t+1} = 0.9s_{i,t}$ . Firm with a performance equal to  $z_t$  do not change in size: if  $p_i = z_t$  then  $s_{i,t+1} = s_{i,t}$ .

Because all firms survive in this model the proportion of high skill firms remains the same in all periods, on average 50%. The market-share weighted average of skill, i.e.,  $\sum_{i=1}^n m_{i,t} u_i$ , changes over time, however. Figure 9 plots how the market-share weighted average of skill changes over time when A)  $\varepsilon_i$  is drawn from a normal distribution with mean 0 and variance 1 or B)  $\varepsilon_i$  is drawn from a t-distribution with mean 0 and 1 degree of freedom. When  $\varepsilon_i$  are drawn from a normal distribution the market-share weighted average of  $u_i$  increases over time. The reason is that high performing firms will grow and increase their market share. A high market-share thus indicates relatively high performance. Moreover, when  $\varepsilon_i$  are drawn from a normal distribution, high performance is an indicator that the firm is likely to have high skill ( $u_i = 1$ ). Eventually, the highest performing firm among the 50 firms will reach a market share close to one. When  $\varepsilon_i$  are drawn from a normal distribution, the highest performing firm is very likely to have high skill.

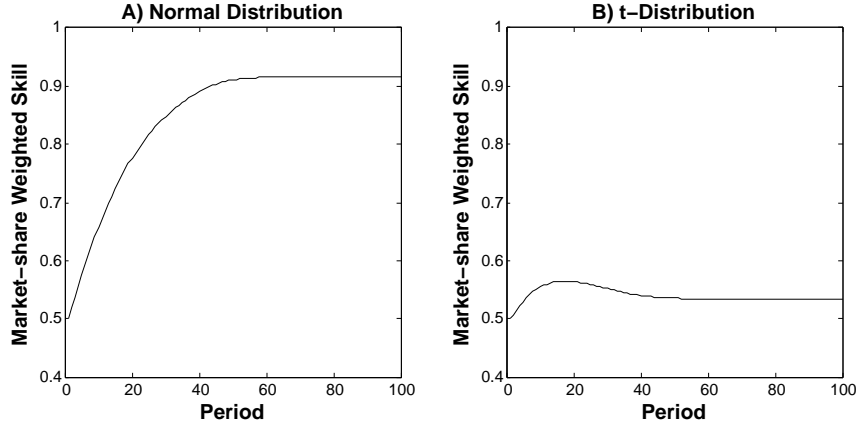


FIGURE 9. How the market-share weighted average of skill changes over time when relatively highly performing firms grow and the noise term is drawn from A) a normal distribution or B) a t-distribution (based on 5000 simulations, each with 50 firms).

By contrast, when  $\varepsilon_i$  are drawn from a t-distribution the market-share weighted average of skill initially increases but then decreases. The reason is that when  $\varepsilon_i$  are drawn from a t-distribution, higher performance is not necessarily an indication that the firm is more likely to have high skill. Firms with moderately high performance are more likely to have high skill than firms with average performance but firms with very high levels of performance are no more likely than firms with average performance to have high skill. Firms with moderately high performance grow initially, but eventually only firms with very high levels of performance grow while others contract. The firm with the highest level of performance eventually reaches a market-share close to one. When  $\varepsilon_i$  are drawn from a t-distribution, a firm with such a high level of performance is not much more likely than an average performing firm to have high skill.

**6.6. Imitation.** Cultural selection can operate via imitation as well as replacement. An agent  $i$  may observe the performance and strategy of another agent  $j$  and switch strategy if agent  $j$  has superior performance and uses a strategy different from what  $i$  is using. A model with these features can also reproduce our basic result.

Specifically, suppose there are  $n$  agents. There is no selection or exit in this model: all agents survive all periods. The model focuses on how agents switch between two ‘strategies’. Each agent can, in each period  $t$ , use one of two possible ‘strategies’:  $u_i = 1$  or  $u_i = 0$ . Thus, we treat the two ‘skill-levels’ as two possible strategies that an agent can adopt. The performance of an agent that uses strategy  $u_i = k$  in period  $t$  is  $p_{i,t} = k + \varepsilon_i$ . Here  $\varepsilon_i$  is a noise term, drawn from a noise distribution  $f_\varepsilon$ . The noise term (and thus performance) remains the same until agent  $i$  changes strategy. Initially, at the start of period 1, 50% of all agents uses strategy  $k = 1$ . Strategy change occurs as follows. In each period  $t$  each agent  $i$  selects at random an agent  $j$  and observes her performance and strategy. If agent  $j$  has higher performance than agent  $i$  has ( $p_{j,t} > p_{i,t}$ ) and  $j$  uses a strategy different from agent  $i$  (i.e.,  $u_j = 1$  while  $u_i = 0$  or  $u_j = 0$  while  $u_i = 1$ ) then agent  $i$  switches strategy to the strategy that  $j$  uses. If agent  $i$  switches strategy her new performance is  $p_i = u_j + \varepsilon_i$  where  $u_j$  is the strategy that  $j$  used and  $\varepsilon_i$  is redrawn, independently, from the noise distribution  $f_\varepsilon$ .

Simulations show that when  $f_\varepsilon$  is a normal distribution then the proportion of agents that uses the strategy  $k = 1$  increases over time (i.e., agents tend to switch to  $k = 1$  over time). When  $f_\varepsilon$  is a t-distribution with 1 degree of freedom, however, the proportion of agents that uses the strategy  $k = 1$  initially increases but eventually decreases. The initial increase occurs because agents that use  $k = 1$  are initially more likely to have high performance. The eventual decrease occurs because after the initial periods of switching to  $k = 1$ , the agents that stick with  $k = 0$  tend to have higher performance than the agents that stick with  $k = 1$ . The reason is that most agents with  $k = 0$  that have low performance are likely to have switched strategy to  $k = 1$ . Eventually, the only agents that stick with  $k = 0$  are those who were lucky, with a high value of  $\varepsilon_i$ .

## 7. ADDITIONAL ROBUSTNESS CHECKS

**7.1. Auto-regressive Performance.** In Section 2, we showed that inefficient selection can happen when performance follows a random walk (Figure 3). The random walk specification assumes that a random draw during the first period of the lifetime of an agent,  $\varepsilon_{i,1}$ , remains relevant during the lifetime of an agent. In some cases it may be more realistic to assume that the impact of  $\varepsilon_{i,t}$  decays over time. This can be modeled by assuming that performance follows an autoregressive structure. Specifically, consider an agent that entered in period  $t$ . Her performance in period  $t$ ,  $p_{i,t}$ , equals her skill,  $u_i$  plus a noise term,  $\varepsilon_{i,t}$ :  $p_{i,t} = u_i + \varepsilon_{i,t}$ . Her performance in period  $t + 1$  equals  $p_{i,t+1} = bp_{i,t} + \varepsilon_{i,t+1}$ . Similarly,  $p_{i,t+2} = bp_{i,t+1} + \varepsilon_{i,t+2}$ , etc. When  $b = 1$ , this specification is

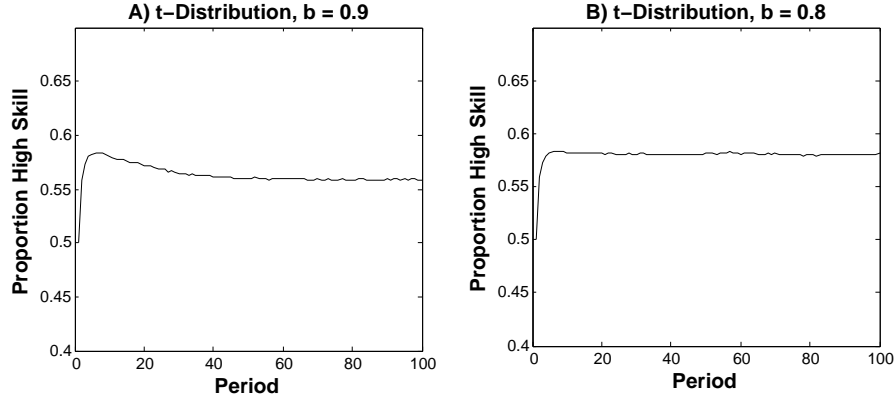


FIGURE 10. How the proportion of high skill agents changes over time when performance follows an autoregressive process,  $p_{i,t+1} = bp_{i,t} + \varepsilon_{i,t+1}$ , and A)  $\varepsilon_{i,t}$  are drawn from a t-distribution with 1 degree of freedom,  $w = 0.35$ , and  $b = 0.9$  and B)  $\varepsilon_{i,t}$  are drawn from a t-distribution with 1 degree of freedom,  $w = 0.35$ , and  $b = 0.8$ . Each graph is based on 10,000 simulations, each with 50 firms, where the  $w$  percent firms with the lowest performances are replaced in every period.

identical to the random walk specification above. Lower values of  $b$  imply lower levels of dependence between performances in consecutive periods (i.e., lower autocorrelation).

Simulations show that the proportion of high skill agents can decline even when  $b$  is lower than 1. To illustrate this, suppose there are 50 agents. Each agent is equally likely to have high ( $u_i = 1$ ) or low ( $u_i = 0$ ) skill. In each period  $t$  the agents with the  $w$  percent lowest performance ( $p_{i,t}$ ) are removed from the population. Each eliminated agent is replaced by a new agent. The probability that a new agent has high skill ( $u_j = 1$ ) is 50%. Figure 10 plots how the proportion of high skill agents changes over time when A)  $\varepsilon_{i,t}$  are drawn from a t-distribution with 1 degree of freedom,  $w = 0.35$ , and  $b = 0.9$  B)  $\varepsilon_{i,t}$  are drawn from a t-distribution with 1 degree of freedom,  $w = 0.35$ , and  $b = 0.8$ . As shown, the proportion of high skill agents eventually declines in the first case, when  $b = 0.9$ , while the proportion of high skill agents initially increases and then reaches a plateau when  $b = 0.8$ .

**7.2. Many Levels of Skill.** Our analysis assumed that there were two possible levels of skill. What if there are many levels of skill or possibly a continuous skill distribution? Simulations show that the basic result can hold even if there is a continuous skill distribution. Suppose, for example, that  $u_i$  is drawn from a normal distribution with mean 0 and variance 1 and remains the same

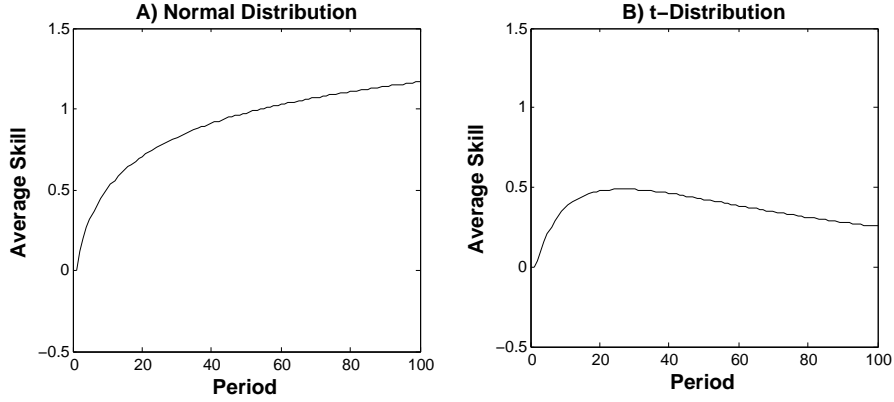


FIGURE 11. How the average level of skill changes over time when skill is drawn from a normal distribution with mean 0 and variance 1 and the noise term is drawn from A) a normal distribution with mean 0 and variance 1 B) a t-distribution with 1 degree of freedom. Each graph is based on 10,000 simulations, each with 50 firms, where the 10% firms with the lowest performances are replaced in every period.

during the lifetime of a firm. The performance of firm  $i$  in period  $t$ ,  $p_{i,t}$ , is equal to the sum of its capability,  $u_i$ , and a noise term,  $\varepsilon_{i,t}$ :  $p_{i,t} = u_i + \varepsilon_i$ . In each period  $t$  the 10 percent lowest performing firms (based on  $p_{i,t}$ ) are removed from the population. Each eliminated firm is replaced by a new firm with performance  $p_{j,t} = u_j + \varepsilon_j$ , where  $u_j$  is drawn from a normal distribution with mean 0 and variance 1 and  $\varepsilon_j$  is drawn from density  $f_\varepsilon$ .  $\varepsilon_j$  remain the same during the lifetime of a firm. Figure 11 shows how the average level of skill changes over time (based on 10,000 simulations each with  $n = 50$  firms). When  $f_\varepsilon$  is a normal distribution with mean 0 and variance 1, the average level of skill increases over time. When  $f_\varepsilon$  is a t-distribution with 1 degree of freedom the average level of skill initially increases but eventually decreases.

The case of a continuous skill distribution is more difficult to handle formally than a binary distribution. What can be demonstrated formally is that whenever the hazard function of the noise distribution is increasing (which is true, for example, for a normal distribution) then average skill increases over time. Formally, suppose that skill (i.e.,  $u_i$ ) is drawn from a distribution with density  $g_u$ , where  $g_u$  can be continuous or discrete. As before, we denote by  $f_\varepsilon$  the density of the noise term and suppose performance equals skill plus the noise term drawn in period 1:  $p_{i,t} = u_i + \varepsilon_{i,1}$ .

**Theorem 5.** *If the hazard function of the distribution of the noise term,  $h_\varepsilon(x)$ , is an increasing function of  $x$ , then average skill in the population increases over time.*

*Proof.* See Appendix E. □

We have not been able to derive necessary or sufficient conditions for when average skill eventually declines. On the basis of simulations, however, we conjecture that average skill eventually decreases if the skill distribution  $g_u$  has an increasing hazard function (such as the normal) and the noise distribution  $f_\varepsilon$  has a decreasing or eventually decreasing hazard function (such as the t-distribution with 1 degree of freedom). By contrast, if the skill distribution  $g_u$  and the noise distribution  $f_\varepsilon$  both have decreasing hazard functions (for example, they are both t-distributions) then average skill increases over time. The general lesson seems to be that average skill can decrease over time when the noise terms are drawn from a distribution with a ‘longer’ tail than the skill distribution (cf. Denrell & Liu (2012)).

## 8. WHEN ARE THE RESULTS RELEVANT?

Our model shows that the proportion of high skill agents eventually declines as a result of selection when four conditions are satisfied:

- (1) The hazard function of the noise distribution is always decreasing (Theorem 2b) or eventually decreasing (Theorem 2c). Theorem 3 also shows that the proportion of high skill agents will converge to 50% as  $t \rightarrow \infty$  when the distribution of the noise term is ‘long-tailed’, a sub-class of heavy-tailed distributions.
- (2) The impact of skill on performance is limited. Theorem 2 assumes that there are only two levels of skill but Figure 11 shows that the average skill also declines over time when skill is drawn from a normal distribution (which is not long-tailed) and the noise term is drawn from a t-distribution (which is long-tailed).
- (3) Performance in a given period strongly depends on performance in earlier periods. Theorem 2 assumes that the level of performance remains the same during the lifetime of an agent, but Figure 3 shows that the proportion of high skill agents may also decline over time when the noise term follows a random walk. Figure 10 shows that we can get the same result when performance follows an autoregressive process,  $p_{i,t+1} = bp_{i,t} + \varepsilon_{i,t+1}$ , as long as  $b$  is sufficiently close to one.
- (4) The threshold level of performance required for survival increases over time, as a result of selection based on relative performance.

In what realistic settings do these technical conditions hold?

Consider first the conditions regarding the noise distribution. The concept of a ‘long-tailed’ distribution is perhaps the easier one to understand intuitively. Informally, a distribution is long-tailed if the probability of an extremely high level of performance for an agent with high skill (the probability that  $1 + \varepsilon_i > c$  when  $c \rightarrow \infty$ ) is about the same as the probability of an extremely high level of performance for an agent with low skill (the probability that  $\varepsilon_i > c$  when  $c \rightarrow \infty$ ). Intuitively, if the error term is drawn from a long-tailed distribution, then the level of skill has almost no impact on the probability of an extreme event. It follows that an extreme event (a very high level of performance) is not informative about the level of skill of the agent.

Several well-known heavy-tailed distributions are long-tailed, including the Pareto distribution, the Log-normal distribution, and the Cauchy distribution (i.e., the t-distribution with one degree of freedom). Moreover, researchers have shown that these distributions fit several important economic outcomes and social indicators. The Pareto distribution and the Log-normal fit the distribution of wealth. (e.g., the wealth of the top 1% population follows a Pareto distribution whereas the wealth of the rest of the population follow a Log-normal distribution, see Levy & Levy (2003)). Stock-market returns fit a t-distribution (Blattberg and Gonedes, 1974). Our model applies in settings where the performance of an agent equals a skill component plus a random draw from these distributions. Our model does not apply in settings where performance is subject to a noise term drawn from a light-tailed distribution such as the normal distribution.

One setting where our results are relevant is firm size. The firm size distribution fits a Log-normal distribution with an upper Pareto tail (Growiec et al., 2008). Moreover, the empirical evidence suggests that random variation, rather than systematic variation in growth rates, account for most of the variance in firm size (Geroski, 2005; Coad, 2007). This suggests that repeated selection based on firm size (which could occur if there are economies of scale, implying that size strongly impacts performance) can lead to a decline in average firm ‘capability’ over time. Our results also have interesting implications for firm growth rates. Evidence suggests that firm growth rates follow a Laplace distribution (Bottazzi et al., 2001; Bottazzi & Secchi, 2003)<sup>3</sup> with a low degree of autocorrelation (Coad, 2007). The Laplace distribution is heavy-tailed distribution which is not ‘long-tailed’. Thus, the result that average ‘capability’ may decline as a result of repeated selection does not hold for this distribution. However, the hazard function for the Laplace distribution is eventually constant (see Table 1) and Theorem 2d implies that, for such a noise distribution, the proportion of high skill agents does not increase over time, as a result of repeated selection, but reaches a plateau. This result may be relevant to explaining the puzzling weak association between

---

<sup>3</sup>Fu et al. (2005), however, argue that the upper-tail is Pareto



productivity and growth (Bottazzi et al., 2001, 2010): repeated selection based on growth will not lead to an increasing proportion of highly productive firms.

The second condition is realistic in settings where variation in skill is limited and unlikely to be responsible for extreme outcomes. Consider trading: it is possible that an individual without skill (e.g., without above average ability to make money in the stock market) might obtain a really high return from trading during one year. Systematic variation in trading ability may exist but explains only a small percentage of the variance in trading results. In other tasks it is inconceivable that low skilled individuals will obtain extremely high outcomes. Consider the 100-meter dash. An unskilled individual who runs 100 meters in 15 seconds, on average, will not, by luck, be able to run below 10 seconds. Many economically relevant tasks are similar: low quality producers will not, by luck, be able to turn out high quality products. Nevertheless, economic outcomes such as profitability, which depend on demand in addition to technical skill, are subject to many uncontrollable factors. For example, demand for cultural products can be very difficult to forecast (Salganik et al., 2006). As a result, it is conceivable that a low-quality producer, who happened to produce what became a fashionable product, would become very profitable.

The third condition, regarding the dependence of current performance on past performance, is relevant in economic and social systems in which there are strong path-dependencies (Merton, 1968; Lynn et al., 2009) such that high past performance makes future high performance more likely. For example, producers with a high market-share may, as a result of network externalities or social influence, be more likely to obtain a high market share in the future. Similarly, individuals who have performed well in their job may, as a result of a self-fulfilling prophecy, be more likely to perform well in the future. Performance in consecutive periods will also be dependent when performance accumulates over time, as in a random walk (Levinthal, 1991; Denrell, 2004). The third condition is less likely to be satisfied in settings where technological or social changes imply that capabilities or resources developed in the past become less applicable or relevant in the future.

The fourth condition - an increasing threshold level of performance required for selection - seems realistic in many economic and social settings where there is repeated selection and survival requires beating other agents. For example, students as well as managers get repeatedly selected for more advanced degrees or positions. Surviving to the next 'stage' often requires beating the other available candidates. Survival in most markets requires offering a product or service that is competitive compared to the offerings of competing firms (Barnett & Hansen, 1996). The fourth condition would not be satisfied if all agents that passed an absolute threshold survived.

When are these four conditions applicable to the selection of firms? An example consists in firms producing a product with success subject to strong network externalities. These firms may differ in their research and development capabilities but their level of performance will strongly depend on whether their product became popular early on and generated a large installed base; an outcome which is related to product quality but also depends on many factors beyond the control of management. If product demand is subject to strong network externalities, performance will be subject to a rich-get-richer dynamics (Arthur, 1989; Barabasi & Albert, 1999) that can generate a heavy-tailed distribution of outcomes (Simon, 1955; Barabasi & Albert, 1999). Over time, firms with poor performance will exit the industry and firms with high performance will grow. Eventually the industry will be dominated by one or a few firms with high market shares. Our model implies that the average research and development capabilities of survivors may decline systematically over time, after an initial increase (compare Figure 9).

When are the four conditions be applicable to selection of individuals within organizations? Selection among individuals in academia provides a possible illustration. Academics are evaluated for jobs, tenure, and chairs based primarily on their research output such as high prestige publications, impact, and citations. Both the number of publications and citations follow heavy-tailed distributions such as the Pareto and the Log-normal (Radicchi et al., 2008). Research on the Matthew Effect suggests that evaluation in academia is noisy and good luck can have a persistent impact because good performance leads to increased attention and resources and improves the chances of high subsequent performance (Merton, 1968). Performance is also persistent because both the number of publications and citations are added up throughout a career. Finally, advancement within schools and to universities of higher status often depends on performance relative to peers. Overall, these observations suggest that it is possible that average research skills decreases over time, as academics are subject to more and more selections.

These illustrations of the possibility that average ‘skill’ declines over time are only relevant if evaluators care about such ‘skill’ rather than performance. For example, hiring committees at universities may want to hire highly cited researchers because they believe citations indicate skill or because they believe having highly cited researchers attracts attention and enhances the reputation of the university. Our result would only be relevant in the first case. Similarly, an investor evaluating a firm may only care about its predicted future performance and not directly about its capability. Even if evaluators only care about performance, however, evaluating skill may be important for predicting future performance. Consider an investor who wants to predict the future performance of a high-technology company that did achieve a dominant position in an

industry that is now subject to technological disruption. Is this firm also likely to achieve dominance in the future? Evaluating this requires an assessment of the capabilities of the firm in developing and marketing new technologies. Was the success of the firm the result of such capabilities - which are likely to remain relevant in the future - or the result of good initial fortune - which is unlikely to be replicated? Similarly, evaluating whether a star individual will continue to perform well in a new job requires an assessment of whether her job performance is due to transferable skills or to idiosyncratic factors (e.g., favorable interpersonal relationships or affiliation with high status firm) that do not apply to the new job (Groysberg, 2010). Finally, one setting in which an evaluator may care about ‘skill’ is the situation we described at the very beginning of the paper, i.e., an outside observer is interested in imitating a high performing firm. Consider a firm with a large installed base. An outside observer would ideally like to imitate all aspects of this firm, including its large installed base. But the large installed base is a consequence of skill and good initial fortune and an imitator may only be able to imitate the skill component.

A possibility that we have not considered is that ‘skill’ evolves over time. Skills may increase as a result of being exposed to competition (Barnett, 2008). If the skills of all firms increase at the same rate our basic result still holds in the sense that an observer who examines firms that have been subject to moderate levels of selection is more likely to identify the most capable firm than an observer who observes firms subject to many periods of selection. It is possible, however, that the skills of high performing firms increase at a higher rate than the skills of poorly performing firms because high performing firms have the resources to invest in capability building. If so, our main result may not hold: the survivors of selection, with high performance, may end up with the highest level of skill. This does not necessarily imply that an outside observer should imitate the survivors of many periods. An outside observer may not be able to imitate the aspects of skill that are the result of investment. Only firms with high performance have the resources to make the necessary investments and develop these skills. As such, these aspects of skills cannot be immediately imitated but are the byproduct of good performance. The relevant aspects of skill for an outside observer might thus be the initial level of skill and for this initial level of skill our main result still holds.

## 9. IMPLICATIONS FOR SELECTION EXPLANATIONS

A selection explanation focuses on changes in the composition of a population instead of changes within the agents making up the population. According to a selection explanation, the reason why most agents in a population have a trait  $x$  rather than  $y$  is not that agents who had  $x$  have switched to  $y$ , but that agents with  $y$  have left the population through a selection process.

The results in this paper challenge the applicability of some seemingly intuitive selection explanations but also suggest that a selection perspective can be extended to novel phenomena. The results challenge the intuition that the survivors of ‘more’ selection are ‘better’ than the survivors of ‘less’ selection. It is not necessarily true that a trait  $x$  that increases expected performance will become more common in a population as selection continues for more periods (Theorem 2). Moreover, it is not necessarily true that the survivors of more ‘intense’ selection, i.e., selection where a larger percentage of the population is removed ( $w$  is larger), have a higher proportion of the trait  $x$  that increases expected performance (Lemma 2 and Theorem 3).

At the same time, our results can also be said to extend the scope of selection explanations because our results show that a selection process can account for non-monotonic effects; effects that may have seemed inconsistent with a selection explanation. For example, consider a researcher who observes that the average level of skill among survivors of selection initially increases but eventually decreases. If this researcher assumes that the survivors of more rounds of selection must have higher expected skill then he or she would need to postulate some additional mechanism to account for the observed decline. Our results show that such additional mechanisms may not be necessary. Below we comment briefly on the implications of these challenges and opportunities for selection explanations used in strategic management and organization theory.

**9.1. Competition and Density Delay.** Our results about more or less intense selection (Lemma 2 and Theorem 3) have implications for discussions about the effect of competition at founding. If only a fixed number of agents can survive, the threshold for survival increases when there is more competition in the sense that there are more firms in the industry (Barnett et al., 2003). Are actors that survive such intense competition better? Barnett et al. (2003) suggest that they likely are, implicitly invoking an assumption that ‘tougher’ selection leads to ‘better survivors’. Our results show that this is not necessarily true. Survivors of tougher selection could in fact be worse. In fact, it is possible that very intense competition can lead to an uninformative selection process. Suppose that  $p_i = u_i + \varepsilon_i$  where  $u_i = 1$  with probability 0.5 and  $u_i = 0$  otherwise. Suppose, further, that higher competition increases the threshold performance,  $c$ , required for survival. Theorem 3 shows that if  $\varepsilon_i$  is drawn from a long-tailed distribution then  $Pr(u_i = 1 | p_i > c) \rightarrow 0.5$  as  $c \rightarrow \infty$ . Formulated differently: very intense selection can be uninformative about skill. This observation also suggests a different interpretation of the finding that density at founding has a persistent negative effect on survival rates (Carroll & Hannan, 1989). This finding has been attributed to a negative effect of competition on capabilities but our results show it can be the result of selection.

**9.2. Obsolescence.** To explain the decline of organizational performance with organizational age for old organizations, researchers have postulated that age or experience has a detrimental effect or that the environment shifts in a way unfavorable to established firms (Hannan et al., 2007; Le Mens et al., 2015a,b). Our results suggest that continued selection can in fact lead to a type of obsolescence: the average skill of the survivors of selection may eventually go down implying that older firms could, as a result of selection rather than adaptation, have lower expected skill than younger firms.

More precisely, if the noise terms are dependent and drawn from a long-tailed distribution, average skill may initially increase then eventually decrease. Thus, continued selection can explain a non-monotonic association between age and organizational skill or ‘capability’. Such a non-monotonic effect has previously been attributed to a combination of selection, generating the initial increase, and environmental drift, resulting in the eventual decline (Hannan et al., 2007). Alternatively, the eventual decrease has been attributed to the negative consequences for capability development of being buffered from competition (Barnett & Hansen, 1996; Barnett, 1997, 2008). Our results show that it is not necessary to postulate environmental drift or capability decline.

This selection-based explanation of obsolescence only predicts a decline in performance in activities that depend on skill and for which the ‘noise term’ is redrawn. The empirical implication is that a decline in performance should only occur in ‘novel’ activities; activities to which previous good fortune does not apply. In activities that continue to draw upon whatever good initial fortune a firm has had, performance should continue to increase. Interestingly, this is broadly consistent with empirical evidence that incumbent firms continue to do well in the absence of radical technological changes (Hill & Rothaermel, 2003). For example, Sorensen & Stuart (2000), in their study of the effect of age on organizational innovativeness, note that the rate of patenting increases with age, suggesting a positive effect of age. As firms age, however, the citations to their new patents by other firms eventually decrease. The usual interpretation of this decrease is that the knowledge of older firms gradually becomes less relevant for others due to environmental drift. An alternative selection explanation, suggested by our results, is that the patenting rate (which increases over time) reflects past good fortune (e.g., to survive, firms need successful innovations on which they subsequently can build, past success also attracts talented employees, etc.). By contrast, the extent to which new patents by the same firm is cited by others reflects innovative ability. Innovative ability ultimately decline because survival during many periods eventually becomes less diagnostic about innovative ability.

More research is needed to develop this idea further and to thoroughly compare it to the existing explanations. Prevailing explanations attribute the decline in average skill to changes within individual units. By contrast, our selection based explanation attributes it to changes in the composition of the population: high skill organizations eventually become more likely to fail in any given period. This explanation is in the same spirit as the random-walk model of Levinthal (1991) although this model focused on the liabilities of newness (higher failure rate for young organizations) and adolescence (the failure rate increases and then decreases with age) but not on the liability of obsolescence.

**9.3. Career Systems.** Many professional organizations such as consulting firms or university departments have an ‘up or out’ career system in which the lowest performers are forced to exit while the others are promoted to the next level. Our results suggest that the individuals promoted to the highest levels can be worse than the individuals who do not reach the highest levels. This occurs when initial good fortune has a long-lasting effect (constant noise terms) and noise can substantially impact the probability of extreme performance (long-tailed noise term distribution). In this scenario the individuals promoted to the highest levels will continue to perform well (because performance depends on past good fortune as well as skill). If they move to a new environment (e.g. a new firm), their expected performance will be lower than those who did not reach the highest levels. Our results imply that this effect can occur due to selection and does not require any change at the individual level.

This analysis offers a novel interpretation of the finding that ‘stars’ hired from another organization perform much worse for the new organization than for their previous employer (Groysberg et al., 2004). This finding has been attributed to laziness resulting from past success or resentment from employees in the new organization (Groysberg et al., 2008). It has also been suggested that the decline may reflect context-dependent skills (Groysberg et al., 2008). The selection explanation is instead that the skills of the ‘stars’, who have survived many rounds of selection, may not be exceptionally high. In fact, the expected skill of stars might be lower than the expected skill of those who have achieve an intermediate level of career success.

## REFERENCES

- Alchian, A. (1950). Uncertainty, evolution, and economic theory. *Journal of Political Economy*, 58(3), 211–221.

- Arthur, B. (1989). Competing technologies, increasing returns, and lock-in by historical events. *The Economic Journal*, 99(394), 116–131.
- Bagnoli, M., & Bergstrom, T. (2005). Log-concave probability and its applications. *Economic Theory*, 26(2), 445–469.
- Barabasi, A., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439), 509–512.
- Barnett, W. P. (1997). The dynamics of competitive intensity. *Administrative Science Quarterly*, 42(1), 128–160.
- Barnett, W. P. (2008). *The Red Queen Among Organizations: How Competitiveness Evolves*. Princeton University Press, Princeton, NJ.
- Barnett, W. P., & Hansen, M. T. (1996). The red queen in organizational evolution. *Strategic Management Journal*, 17(S1), 139–157.
- Barnett, W. P., Swanson, A. N., & Sorenson, O. (2003). Asymmetric selection among organizations. *Industrial and Corporate Change*, 12(4), 673–695.
- Bottazzi, G., Dosi, G., Jacoby, N., Secchi, A., & Tamagni, F. (2010). Corporate performances and market selection: some comparative evidence. *Industrial and Corporate Change*, 19(6), 1953–1996.
- Bottazzi, G., Dosi, G., Lippi, M., Pammolli, F., & Riccaboni, M. (2001). Innovation and corporate growth in the evolution of the drug industry. *International Journal of Industrial Organization*, 19(7), 1161–1187.
- Bottazzi, G., & Secchi, A. (2003). Why are distributions of firm growth rates tent-shaped? *Economics Letters*, 80(3), 415–420.
- Carroll, G., & Hannan, M. (1989). Density delay in the evolution of organizational populations: a model and five empirical tests. *Administrative Science Quarterly*, 34(3), 411–430.
- Carroll, G. R., & Harrison, J. R. (1994). On the historical efficiency of competition between organizational populations. *American Journal of Sociology*, 100(3), 720–749.
- Coad, A. (2007). A Closer Look at Serial Growth Rate Correlation. *Review of Industrial Organization*, 31(1), 69–82.
- Cohen, D. (1966). Optimizing reproduction in a randomly varying environment. *Journal of Theoretical Biology*, 12, 119–129.
- Cvijovic, I., Good, B. H., Jerison, E. R., & Desai, M. M. (2015). Fate of a mutation in a fluctuating environment. *Proceedings of the National Academy of Sciences*, 112(36), 5021–5028.

- Denrell, J. (2004). Random walks and sustained competitive advantage. *Management Science*, 50(7), 922–934.
- Denrell, J., & Liu, C. (2012). Top performers are not the most impressive when extreme performance indicates unreliability. *Proceedings of the National Academy of Sciences of the United States of America*, 109(24), 9331–9336.
- Denrell, J., & March, J. G. (2001). Adaptation as information restriction: The hot stove effect. *Organization Science*, 12(5), 523–538.
- Elster, J. (1979). *Ulysses and the Sirens: Studies in rationality and irrationality*. Cambridge University Press, Cambridge, U.K.
- Foss, S., Korshunov, D., & Zachary, S. (2013). *An Introduction to Heavy-Tailed and Subexponential Distributions*. Springer Series in Operations Research and Financial Engineering.
- Fu, D., Pammolli, F., Buldyrev, S. V., Riccaboni, M., Matia, K., Yamasaki, K., & Stanley, H. E. (2005). The growth of business firms: Theoretical framework and empirical evidence. *Proceedings of the National Academy of Sciences of the United States of America*, 102(52), 18801–18806.
- Geroski, P. A. (2005). Understanding the implications of empirical work on corporate growth rates. *Managerial and Decision Economics*, 26(2), 129–138.
- Gimeno, J., Folta, T. B., Cooper, A. C., & Woo, C. Y. (1997). Survival of the fittest? Entrepreneurial human capital and the persistence of underperforming firms. *Administrative Science Quarterly*, 42(4), 750–783.
- Glaser, R. E. (1980). Bathtub and related failure rate characterizations. *Journal of the American Statistical Association*, 75(371), 667–672.
- Growiec, J., Pammolli, F., Riccaboni, M., & Stanley, H. E. (2008). On the size distribution of business firms. *Economics Letters*, 98(2), 207–212.
- Groysberg, B. (2010). *Chasing stars: The myth of talent and the portability of performance*. Princeton University Press.
- Groysberg, B., Nanda, A., & Nohria, N. (2004). The risky business of hiring stars. *Harvard Business Review*, 82(5), 92–101.
- Groysberg, B., Sant, L., & Abrahams, R. (2008). When stars migrate, do they still perform like stars? *MIT Sloan Management Review*, 50(1), 41–46.
- Hannan, M. T., Polos, L., & Carroll, G. R. (2007). *Logics of Organization Theory: Audiences, Codes, and Ecologies*. Princeton, NJ: Princeton University Press.
- Hill, W., Charles, & Rothaermel, F. (2003). The Performance of incumbent firms in the face of radical technological innovation. *Academy of Management Review*, 28(2), 257–274.



- Hodgson, G. M., & Knudsen, T. (2010). *Darwin's conjecture: The search for general principles of social and economic evolution*. University of Chicago Press, Chicago, IL.
- Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems*. Ann Arbor, MI: University of Michigan Press. Second edition, 1992.
- Jovanovic, B. (1982). Selection and the evolution of industry. *Econometrica*, 50, 649–670.
- Karlin, S., & Rubin, H. (1956). Distributions possessing a monotone likelihood ratio. *Journal of the American Statistical Association*, 63, 637–643.
- Le Mens, G., Hannan, M., & Pólos, L. (2015a). Age-related structural inertia: A distance-based approach. *Organization Science*, 26(3), 756–773.
- Le Mens, G., Hannan, M., & Pólos, L. (2015b). Organizational obsolescence, drifting tastes, and age dependence in organizational life chances. *Organization Science*, 26(2), 550–570.
- Le Mens, G., Hannan, M. T., & Pólos, L. (2011). Founding conditions, learning, and organizational life chances: age dependence revisited. *Administrative Science Quarterly*, 56(1), 95–126.
- Levins, R. (1968). *Evolution in Changing Environments*. Princeton University Press, Princeton NJ.
- Levinthal, D., & March, J. G. (1981). A model of adaptive organizational search. *Journal of Economic Behavior and Organization*, 2(4), 307–333.
- Levinthal, D., & Marino, A. (2015). Three facets of organizational adaptation: Selection, variety, and plasticity. *Organization Science*, 26(3), 743–755.
- Levinthal, D., & Posen, H. (2007). Myopia of selection: Does organizational adaptation limit the efficacy of population selection? *Administrative Science Quarterly*, 52(4), 586–620.
- Levinthal, D. A. (1991). Random walks and organizational mortality. *Administrative Science Quarterly*, 36(3), 397–420.
- Levinthal, D. A. (1997). Adaptation on rugged landscapes. *Management Science*, 43(7), 934–950.
- Levy, M., & Levy, H. (2003). Investment talent and the pareto wealth distribution: Theoretical and experimental analysis. *Review of Economics and Statistics*, 85(3), 709–725.
- Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. New York, NY: New York, NY: Oxford University Press.
- Lynn, F. B., Podolny, J. M., & Tao, L. (2009). A sociological (de) construction of the relationship between status and quality. *American Journal of Sociology*, 115(3), 755–804.
- Maynard Smith, J. (1982). *Evolution and the Theory of Games*. Cambridge University Press.
- Merton, R. K. (1968). The Matthew effect in science. *Science*, 159(3810), 56–63.

- Meyer, M. W., & Zucker, L. (1989). *Permanently failing organizations*. Sage Publications, London, UK.
- Nair, J., Wierman, A., & Zwart, B. (2013). The Fundamentals of Heavy-tails: Properties, emergence, and identification. *Proceedings of the ACM SIGMETRICS*, 41(1).
- Nelson, R., & Winter, S. (2002). Evolutionary theorizing in economics. *The Journal of Economic Perspectives*, 16(2), 23–46.
- Nelson, R. R., & Winter, S. G. (1982). *An evolutionary theory of economic change*. Cambridge, MA: Harvard University Press.
- Perrow, C. (2002). *Organizing America: Wealth, Power, and the Origins of Corporate Capitalism*. Princeton, NJ: Princeton University Press.
- Radicchi, F., Fortunato, S., & Castellano, C. (2008). Universality of citation distributions: Toward an objective measure of scientific impact. *Proceedings of the National Academy of Sciences*, 105(45), 17268–17272.
- Ross, S. (2000). *An Introduction to Probability Models, Seventh Edition*. Harcourt Academic Press, San Diego, CA.
- Salganik, M., Dodds, P., & Watts, D. J. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, 311(5762), 854–856.
- Simon, H. (1955). On a class of skew distribution functions. *Biometrika*, 42(3-4), 425–440.
- Sorensen, J., & Stuart, T. E. (2000). Aging, obsolescence, and organizational innovation. *Administrative Science Quarterly*, 45(1), 81–112.
- Thomas, E. (1971). Sufficient conditions for monotone hazard rate: An application to latency-probability curves. *Journal of Mathematical Psychology*, 8, 303–332.
- Weibull, J. W., Mattsson, L. G., & Voorneveld, M. (2007). Better may be worse: Some monotonicity results and paradoxes in discrete choice under uncertainty. *Theory and Decision*, 63(2), 121–151.
- Winter, S. (1964). Economic 'natural selection' and the theory of the firm. *Yale Economic Essays*, 4(1), 225–272.
- Winter, S. (1971). Satisficing, selection and the innovating remnant. *Quarterly Journal of Economics*, 85(2), 237–261.
- Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, 16, 97–159.
- Yoshimura, J., & Clark, C. W. (1993). *Adaptation in stochastic environments*. Springer-Verlaag, New York, NY.

## APPENDIX A. PROOF OF THEOREM 1

We start by proving the following lemma that applies to the general model setup described in Section 4.1.

**Lemma 1.** *In the setting described in section 4.1, the proportion of high skill agents,  $\pi_t$ , increases from period  $t$  to period  $t + 1$  whenever the probability that a low skill agent exits during period  $t + 1$  is higher than the probability that a high skill agent exits during period  $t + 1$ .*

*Proof.* Let  $\Pr_{H,t+1}$  be the probability that a high skill agent survives during period  $t + 1$ . Similarly, let  $\Pr_{L,t+1}$  be the probability that a low skill agent survives during period  $t + 1$ . We can express the proportion of high skill agents at the end of period  $t + 1$  as a function of the proportion of high skill agents at the end of period  $t$  and the failure probabilities of high skill and low skill agents in period  $t + 1$ . We have:

$$(2) \quad \pi_{t+1} = \frac{\pi_t \Pr_{H,t+1}}{\pi_t \Pr_{H,t+1} + (1 - \pi_t) \Pr_{L,t+1}} = \pi_t \frac{\Pr_{H,t+1}}{\pi_t \Pr_{H,t+1} + (1 - \pi_t) \Pr_{L,t+1}},$$

If  $\Pr_{H,t+1} > \Pr_{L,t+1}$ , then  $\Pr_{H,t+1} > \pi_t \Pr_{H,t+1} + (1 - \pi_t) \Pr_{L,t+1}$  and  $\pi_{t+1} > \pi_t$ .  $\square$

To prove Theorem 1 by application of this lemma, it is enough to show that in every period, the probability that a low skill agent exits is higher than the probability that a high skill agent exits.

Let  $t \geq 1$ . Let  $p_t^*$  be the maximum level of performance among the agents that failed during period  $t$ . The probability that a low skill agent exits in period  $t$  is  $P(0 + \varepsilon_{i,t} \leq p_t^*)$ . The probability that a high skill agent exits in period  $t$  is  $P(1 + \varepsilon_{i,t} \leq p_t^*)$ . Clearly,  $P(0 + \varepsilon_{i,t} \leq p_t^*) > P(1 + \varepsilon_{i,t} \leq p_t^*)$ . Therefore, the probability that a low skill agent exits is higher than the probability that a high skill agent exits in period  $t$ .

We have shown that in all periods, the exit probability of a high skill agent is higher than the exit probability of a low skill agent. Lemma 1 implies that  $\pi_t$  increases with  $t$  for all  $t$ .  $\square$

## APPENDIX B. PROOF OF THEOREM 2

Here we characterize the noise distributions for which  $\pi_t$  monotonically increases or monotonically decreases with  $t$ . We begin by analyzing how  $P(u_i = 1 \mid p_i > c)$  varies with  $c$ .

Let  $F_\varepsilon(x)$  denote the cumulative density function of the noise term and let  $f_\varepsilon(x)$  denote the corresponding density function. We assume that the support of  $f_\varepsilon$  is of the form  $(a, +\infty)$  with  $a \in \{-\infty, \mathbb{R}\}$  (for all  $x \in (a, +\infty)$ ,  $f_\varepsilon(x) > 0$  and  $f_\varepsilon(x) = 0$  elsewhere). Because  $p_{i,t} = 0 + \varepsilon_{i,1} = \varepsilon_{i,1}$  for low skill actors, the probability density function of performance for low skill actors is  $f_0(x) = f_\varepsilon(x)$ , with support  $(a, \infty)$ , and the cumulative density function is  $F_0(x) = F_\varepsilon(x)$ . For high skill actors,

$p_{i,t} = 1 + \varepsilon_{i,1}$ . Therefore, for high skill actors, the density of performance is  $f_1(x) = f_\varepsilon(x - 1)$ , with support  $(a + 1, \infty)$ , and the cumulative density function is  $F_1(x) = F_\varepsilon(x - 1)$ . Let  $S_0(x)$  be the survival function for actors with low skill, i.e.,  $S_0(x) = 1 - F_0(x)$ , and let  $S_1(x)$  be the survival function for actors with high skill, i.e.,  $S_1(x) = 1 - F_1(x - 1)$ . Finally, let  $h_0(x) = f_0(x)/S_0(x)$  be the hazard function for low skill agents and  $h_1(x) = f_1(x)/S_1(x)$  be the hazard function for high skill agents.

**Lemma 2** (Skill and intensity of selection). *In the setting of Theorem 2 (noise terms drawn at time of entry from the density  $f_\varepsilon(x)$  and no replacement), we have:*

- i) *if for all  $c > a + 1$ ,  $h_1(c) < h_0(c)$  then  $P(u_i = 1 \mid p_i > c)$  is an increasing function of  $c$ .*
- ii) *if for all  $c > a + 1$ ,  $h_1(c) > h_0(c)$ , then  $P(u_i = 1 \mid p_i > c)$  is a decreasing function of  $c$ .*
- iii) *if for all  $c > a + 1$ ,  $h_1(c) = h_0(c)$ , then  $P(u_i = 1 \mid p_i > c)$  is a constant.*

*Proof.* Note that both  $S_0(c)$  and  $S_1(c)$  are decreasing in  $c$  (as are all survival functions). Using Bayes rule, the proportion of high skill agents is given by

$$(3) \quad P(u_i = 1 \mid p_i > c) = \frac{P(p_i > c \mid u_i = 1)\pi_0}{P(p_i > c \mid u_i = 1)\pi_0 + P(p_i > c \mid u_i = 0)(1 - \pi_0)}.$$

Because  $P(p_i > c \mid u_i = 1) = S_1(c)$ ,  $P(p_i > c \mid u_i = 0) = S_0(c)$ , and  $\pi_0 = 0.5$  we get

$$(4) \quad P(u_i = 1 \mid p_i > c) = \frac{S_1(c)0.5}{S_1(c)0.5 + S_0(c)(1 - 0.5)} = \frac{1}{1 + \frac{S_0(c)}{S_1(c)}}.$$

From equation 4, it is clear that  $P(u_i = 1 \mid p_i > c)$  is an increasing function of  $c$  whenever  $S_0(c)/S_1(c)$  is a decreasing function of  $c$ . Moreover,

$$\frac{dS_0(c)/S_1(c)}{dc} = \frac{-f_0(c)S_1(c) + f_1(c)S_0(c)}{S_1(c)^2},$$

which is negative when  $f_1(c)S_0(c) < f_0(c)S_1(c)$ , or,  $f_1(c)/S_1(c) < f_0(c)/S_0(c)$ , i.e., when  $h_1(c) < h_0(c)$ . Similarly,  $P(u_i = 1 \mid p_i > c)$  is decreasing function of  $c$  when  $h_1(c) > h_0(c)$  and  $P(u_i = 1 \mid p_i > c)$  is a constant when  $h_1(c) = h_0(c)$ . Claims i) to iii) follow trivially.  $\square$

Lemma 2 shows that the shape of the hazard function determines whether  $P(u_i = 1 \mid p_i > c)$  increases, decreases or is constant in  $c$ . If the hazard function of the noise term distribution is increasing for all values, implying that  $h_1(c) < h_0(c)$  (because  $h_1(c) = h_\varepsilon(c - 1)$  and  $h_0(c) = h_\varepsilon(c)$ ) then  $P(u_i = 1 \mid p_i > c)$  is increasing in  $c$ . If the hazard function is decreasing, implying that  $h_1(c) > h_0(c)$ ,  $P(u_i = 1 \mid p_i > c)$  is decreasing in  $c$ . If the hazard function initially increases but eventually decreases with  $c$  then  $P(u_i = 1 \mid p_i > c)$  may initially increase and then decrease with

c. To determine whether  $P(u_i = 1 \mid p_i > c)$  increases with  $c$  it is thus enough to know whether the hazard function is increasing or decreasing.

Whether the hazard function is strictly increasing or strictly decreasing can be determined from the shape of the density function as follows:

**Lemma 3.** *Let  $g(x)$  be a differentiable density function such that  $g(x) \rightarrow 0$  as  $G(x) \rightarrow 1$ . If  $\ln g(x)$  is a concave (convex) function of  $x$ , then  $h(x) = g(x)/(1 - G(x))$  is an increasing (decreasing) function of  $x$ .*

*Proof.* See Thomas (1971). □

Using this criterion it can be shown that the hazard function of the normal distribution is increasing (see, Luce (1986), p. 16-17). It follows that  $P(u_i = 1 \mid p_i > c)$  is an increasing function of  $c$  when the noise term is drawn from a normal distribution.

Next, we prove that the threshold for survival increases over time.

**Lemma 4.** *Let  $p_t^*$  be the maximum level of performance among the agents that fail in period  $t$ . In the setting of Theorem 2,  $p_t^*$  increases with  $t$  and  $\lim_{t \rightarrow \infty} p_t^* = \infty$ .*

*Proof.* Because the noise distribution is continuous, the probability that two or more agents have the same level of performance is 0. It follows that all the survivors of selection in period  $t$  have a performance higher than  $p_t^*$ . Because some positive fraction of these survivors from period  $t$  are eliminated in period  $t + 1$ ,  $p_{t+1}^* > p_t^*$ . Hence,  $p_t^*$  increases with  $t$ .

Now we prove that  $p_t^*$  converges toward  $\infty$ . Suppose that  $(p_t^*)_{t \geq 1}$  is bounded by  $k < \infty$ . Let  $\gamma$  denote the initial proportion of agents with performance at least as high as  $k$ :  $\gamma = P(p_{i,1} \geq k)$ . Note that  $\gamma > 0$  because the support of  $f_\varepsilon$  is unbounded on the right. All agents with performance at least as high as  $k$  survive infinitely many periods because their performance is higher than all the performance thresholds (we showed that the sequence of performance thresholds is increasing and we assumed it was bounded). But at the same time, after  $t$  periods the proportion of survivors is  $(1 - w)^t$ . This converges toward 0 when  $t$  becomes large. This fact is incompatible with the fact that at least  $\gamma > 0$  agents survive all periods. Therefore, it cannot be the case that  $p_t^*$  is bounded. Thus, necessarily,  $\lim_{t \rightarrow \infty} p_t^* = \infty$ . □

By combining Lemmas 2 and 4 we can characterize the noise distributions for which  $\pi_t$  always increases with  $t$  and the noise distributions for which  $\pi_t$  eventually decreases with  $t$ . Note that in the setting of the theorem,  $\pi_t = P(u_i = 1 \mid p_i > p_t^*)$ .

*i)* Suppose  $t = 1$ . The proportion of high skill agents who survive the first period,  $P(1 + \varepsilon_i > p_1^*)$  is larger than the proportion of low skill agents who survive the first period,  $P(\varepsilon_i > p_1^*)$ . Hence,  $\pi_1 > 0.5$ .

*ii a)* Suppose the noise distribution has a strictly increasing hazard function  $h_\varepsilon$ . Lemma 2i implies that  $P(u_i = 1 \mid p_i > p_t^*)$  increases with  $p_t^*$ . Lemma 4 implies that  $p_t^*$  increases with  $t$ . By composition, this implies that  $P(u_i = 1 \mid p_i > p_t^*) = \pi_t$  increases with  $t$ .

*ii b)* Suppose  $h_\varepsilon$  is a decreasing hazard function. According to Lemma 4,  $p_t^*$  increases with  $t$  and converges toward  $\infty$ . Let  $t^a$  be the first period in which  $p_t^* > a + 1$  ( $a$  is the lower bound of the support of  $f_\varepsilon$ ). It follows from Lemma 2ii that when  $t > t^a$ ,  $\pi_t$  decreases with  $t$ .

*ii c)* Suppose  $h_\varepsilon(c - 1) < h_\varepsilon(c)$  for all  $c < c^*$  and  $h_\varepsilon(c - 1) > h_\varepsilon(c)$  for all  $c > c^*$ . Let  $t^m$  be the first period in which  $p_{t^m}^* > c^*$  (we can be sure that  $t^m$  exists because  $\lim_{t \rightarrow \infty} p_t^* = \infty$ , as implied by Lemma 4). Let  $t^a$  be the first period in which  $p_t^* > a + 1$  ( $a$  is the lower bound of the support of  $f_\varepsilon$ ). It follows from Lemma 2ii that when  $t > \max(t^m, t^a)$ ,  $\pi_t$  decreases with  $t$ .

*ii d)* Suppose  $h_\varepsilon(c - 1) = h_\varepsilon(c)$  for all  $c > c^*$ . Let  $t^m$  be the first period in which  $p_{t^m}^* > c^*$  and  $t^a$  be the first period in which  $p_t^* > a + 1$ . It follows from Lemma 2iii that when  $t > \max(t^m, t^a)$ ,  $\pi_t$  remains constant.

#### APPENDIX C. PROOF OF THEOREM 3

*i)* According to equation 4, the proportion of high skill agents can be written in terms of the survival functions, and, in turn, in terms of the cumulative distribution function of the noise term:

$$P(u_i = 1 \mid p_i > c) = \frac{S_1(c)0.5}{S_1(c)0.5 + S_0(c)0.5} = \frac{1 - F_\varepsilon(c - 1)}{1 - F_\varepsilon(c - 1) + 1 - F_\varepsilon(c)} = \frac{1}{1 + \frac{1 - F_\varepsilon(c)}{1 - F_\varepsilon(c - 1)}}.$$

Therefore  $\lim_{c \rightarrow \infty} P(u_i = 1 \mid p_i > c) = 0.5$  if and only if  $\lim_{c \rightarrow \infty} \frac{1 - F_\varepsilon(c)}{1 - F_\varepsilon(c - 1)} = 1$ . This latter condition is exactly the necessary and sufficient condition for a distribution to be long-tailed.

*ii)* This follows from *i)* and Lemma 4. □

#### APPENDIX D. PROOF OF THEOREM 4

To prove Theorem 4 note first that, after selection and replacement in period  $t$ ,

$$\pi_t = (1 - w)P(u_i = 1 \mid p_i > p_t^*) + w0.5.$$

The first term represents the proportion of high skill agents among the agents that survived selection: their performance must all be above the threshold for survival in period  $t$ . The second term is the proportion of high skill agents among the new agents, which is 0.5. Because the change in  $\pi_t$  only depends on  $P(u_i = 1 \mid p_i > p_t^*)$  (all other elements of the formula remain constant over time), the conclusion of Theorem 2 continues to hold as long the equivalent of Lemma 4, i.e.,  $p_t^*$  is increasing without bound, continues to hold when there is replacement.

The proof that  $p_t^*$  is increasing follows almost the same reasoning as the proof of Lemma 4. All the survivors of selection in period  $t$  have a performance higher than  $p_t^*$ . Some positive fraction of these survivors from period  $t$  are eliminated in period  $t + 1$ , because, whenever  $p_t^*$  is finite, there is a positive probability that an entrant  $i$  in period  $t + 1$  will have a performance higher than a survivor  $j$  in period  $t$ , i.e.,  $P(p_{i,1} > p_{j,1} \mid p_{j,1} > p_t^*) > 0$ . Hence,  $p_t^*$  increases with  $t$ .

Now we prove that  $p_t^*$  converges toward  $\infty$ . Suppose that  $(p_t^*)_{t \geq 1}$  is bounded by  $k < \infty$ . Let  $\gamma$  denote the proportion of agents with a performance at least as high as  $k$ :  $\gamma = P(p_{i,1} \geq k)$ . Note that  $\gamma > 0$  because the support of  $f_\varepsilon$  is unbounded on the right. All agents with performance at least as high as  $k$  survive infinitely many periods. Initially, the proportion of agents with performance of at least  $k$  is  $\gamma$ . The probability that an entrant has performance of at least  $k$  is also  $\gamma$ . The fraction of entrants is  $w$  in each period. Over time, the proportion of agents with a performance of at least  $k$  becomes higher  $(1 - w)$ . To see why, suppose that in period 1,  $\gamma \leq 1 - w$ . In period 2, we have  $w$  entrants that are replacing some of the agents with performance below  $k$ . Among the entrants, the proportion of agents with performance of at least  $k$  is equal to  $\gamma$ . So at the end of period 2, the proportion of agents with performance of at least  $k$  is equal to  $\gamma + \gamma w$ . If this proportion is higher than  $1 - w$ , we have proved what we wanted. If this proportion lower than or equal to  $1 - w$ , the same dynamics applies to period the following periods. If at the end of period  $t - 1$ , the proportion of agents with performance of at least  $k$  is lower than  $1 - w$ , then this proportion at the end of period  $t$  is  $\gamma + (t - 1)\gamma w$ . This dynamic process necessarily implies that at some point, the proportion of agents with performance of at least  $k$  becomes higher than  $1 - w$ . That means that the proportion of replaced agents becomes lower than  $1 - w$ , which is incompatible with the definition of  $w$ . Therefore, it cannot be the case that  $p_t^*$  is bounded. Thus, necessarily,  $\lim_{t \rightarrow \infty} p_t^* = \infty$ .

## APPENDIX E. PROOF OF THEOREM 5

To prove Theorem 5 it is enough to prove that that the equivalent of Lemma 2i holds, for any skill distribution, whenever the hazard function  $h_\varepsilon$  is increasing:

**Lemma 5.** *In the setting of Theorem 5, if  $h_\varepsilon$  is increasing, then  $E[u_i \mid p_i > c]$  is increasing in  $c$ .*

*Proof.* We denote by  $P$  the performance and by  $U$  the skill of a randomly chosen agent. Consider an agent  $i$  and its performance  $p_i$ . The survival function, conditional on skill, is defined as  $S(c \mid U = u_i) = P(p_i > c \mid U = u_i) = 1 - F_\varepsilon(c - u_i)$ .

Let  $c_1$  and  $c_2$  be performance thresholds such as  $c_2 > c_1$ . We want to show  $E[u \mid p_i > c_2] > E[u \mid p_i > c_1]$ . To do so, we will write  $E[u \mid p_i > c_2]$  as the sum of  $E[u \mid p_i > c_1]$  and a positive term.

Assuming a performance threshold of  $c_1$ , we can use Bayes rule to write the skill distribution for the survivors:

$$(5) \quad g_1(u \mid p_i > c_1) = \frac{P(p_i > c_1 \mid U = u)g_u(u)}{\int_{-\infty}^{\infty} P(p_i > c_1 \mid U = \nu)g_u(\nu)d\nu}.$$

Because  $P(p_i > c_1 \mid U = u) = S(c_1 \mid U = u)$  this can be written as

$$(6) \quad g_1(u \mid p_i > c_1) = \frac{S(c_1 \mid U = u)g_u(u)}{\int_{-\infty}^{\infty} S(c_1 \mid U = \nu)g_u(\nu)d\nu}.$$

Similarly if the performance threshold is  $c_2$ , we can write

$$(7) \quad g_2(u \mid p_i > c_2) = \frac{S(c_2 \mid U = u)g_u(u)}{\int_{-\infty}^{\infty} S(c_2 \mid U = \nu)g_u(\nu)d\nu},$$

$$(8) \quad = \frac{w(u)g_1(u \mid p_i > c_1)}{\int_{-\infty}^{\infty} w(\nu)g_1(\nu \mid p_i > c_1)d\nu},$$

where  $w(u) = S(c_2 \mid U = u)/S(c_1 \mid U = u)$ . We get

$$(9) \quad E[u \mid p_i > c_2] = \frac{\int_{-\infty}^{\infty} uw(u)g_1(u \mid p_i > c_1)du}{\int_{-\infty}^{\infty} w(u)g_1(u \mid p_i > c_1)du} = \frac{E_1[w(u)u]}{E_1[w(u)]},$$

where  $E_1$  denotes that the expectation is with respect to the random variable with density  $g_1(u \mid p_i > c_1)$ . This can be expressed as

$$(10) \quad E[u \mid p_i > c_2] = \frac{E_1[w(u)]E_1[u] + Cov_1[w(u), u]}{E_1[w(u)]} = E[u \mid p_i > c_1] + \frac{Cov_1[w(u), u]}{E_1[w(u)]},$$



where  $Cov_1$  denotes a covariance with respect to density  $g_1(u \mid p_i > c_1)$ . Then,

$$(11) \quad E[u \mid p_i > c_2] = E[u \mid p_i > c_1] + \frac{Cov_1[\frac{S(c_2|U=u)}{S(c_1|U=u)}, u]}{E_1[\frac{S(c_2|U=u)}{S(c_1|U=u)}]}.$$

Whenever  $S(c_2 \mid U = u)/S(c_1 \mid U = u)$  is a strictly increasing function of  $u$ , then the covariance term is positive (Ross (2000), p. 626), and thus  $E[u \mid p_i > c_2]$  is larger than  $E[u \mid p_i > c_1]$ . Moreover,  $S(c_2 \mid U = u)/S(c_1 \mid U = u)$  is a strictly increasing function of  $u$  iff

$$(12) \quad \frac{d}{du} \frac{S(c_2 \mid U = u)}{S(c_1 \mid U = u)} = \frac{f_\varepsilon(c_2)S(c_1 \mid U = u) - f_\varepsilon(c_1)S(c_2 \mid U = u)}{S(c_1 \mid U = u)^2} > 0,$$

This occurs when  $f_\varepsilon(c_1)S(c_2 \mid U = u) < f_\varepsilon(c_2)S(c_1 \mid U = u)$ , i.e., when  $h_\varepsilon(c_2) = f_\varepsilon(c_1)/S(c_1 \mid U = u)$  is smaller than  $h_\varepsilon(c_2) = f_\varepsilon(c_2)/S(c_2 \mid U = u)$ . Since  $c_1 < c_2$  and the hazard function,  $h_\varepsilon$ , is by assumption increasing, this condition is satisfied. Therefore,  $Cov_1[\frac{S(c_2|U=u)}{S(c_1|U=u)}, u] > 0$ . Equation 11 in turns implies that  $E[u \mid p_i > c_2] > E[u \mid p_i > c_1]$ .  $\square$